

GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 29, 2006, 10:41:28 ; Search time 935 Seconds  
(without alignments)  
5100.554 Million cell updates/sec

Title: US-10-718-311-7

Perfect score: 684  
Sequence: 1 atgggttcctctgtcatttc.....cggcgtaaccgttgactaa 684

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5244920 seqs, 3486124231 residues

Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_8.\*

- 1: geneseqn1980s.\*
- 2: geneseqn1990s.\*
- 3: geneseqn2000s.\*
- 4: geneseqn2001as.\*
- 5: geneseqn2001bs.\*
- 6: geneseqn2002as.\*
- 7: geneseqn2002bs.\*
- 8: geneseqn2003as.\*
- 9: geneseqn2003bs.\*
- 10: geneseqn2003cs.\*
- 11: geneseqn2003ds.\*
- 12: geneseqn2004as.\*
- 13: geneseqn2004bs.\*
- 14: geneseqn2005s.\*
- 15: geneseqn2006s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	684	100.0	684	6	ABA91838 Chloropla
2	684	100.0	684	6	Abk47712 DNA seque
3	684	100.0	684	10	Adc25988 Tomato/Es
4	684	100.0	684	14	Adv91822 Tomato ru
5	501	73.2	3452	14	Adv91636 C. paraps
6	499.4	73.0	1371	14	Adv91643 C. paraps
7	498	72.8	498	14	Adv91618 Escherich
8	498	72.8	854	4	Aad06956 ubiC gene
9	498	72.8	6641	4	Aad06957 pME2 comp
10	496.4	72.6	2000	2	Aat29820 E. coli u
11	495	72.4	495	2	Aaq92409 E. coli ch
12	495	72.4	495	6	ABA91837 Escherich
13	495	72.4	495	6	Abk47709 DNA seque
14	495	72.4	495	10	Adc25984 Escherich
15	472.8	69.1	1207	5	Aas90335 DNA encod
16	472.8	69.1	1318	2	Aaq67671 E.coli ub
17	282	41.2	645	11	ACH95439 Klebsiell
18	145.6	21.3	492	10	ADC76164 DNA homol

ALIGNMENTS

RESULT 1

ABA91838

ID ABA91838 standard; DNA; 684 BP.

XX ABA91838;

XX 29-AUG-2003 (revised)

DT 15-MAY-2002 (first entry)

XX Chloroplast transit peptide-chorismate pyruvate lyase fusion gene.

XX Chloroplast transit peptide; tomato; Rubisco; plant;

KW ribulose-1,5-bisphosphate carboxylase; chorismate pyruvate lyase; CPL;

KW enzyme; p-hydroxybenzoic acid; transgenic plant; gene; ds.

XX Lycopersicon esculentum.

OS Escherichia coli.

OS Chimeric.

XX Key Location/Qualifiers

FT CDS 1..495

FT /product= "chloroplast-targeted CPL fusion"

FT transit\_peptide 1..186

FT /tag= a

FT /tag= b

FT /note= "tomato Rubisco chloroplast targeting sequence"

FT mat\_peptide 187..492

FT /tag= c

FT /note= "E. coli chorismate pyruvate lyase"

PN WO200194607-A2.

XX 13-DEC-2001.

XX 22-MAY-2001; 2001WO-US016661.

XX 02-JUN-2000; 2000US-0209854P.

XX (DUPO ) DU PONT DE NEMOURS & CO E I.

XX Meyer K, Van Dyk DE, Viitanen PV;

```
XX WPI; 2002-226795/28.
DR P-PSDB; AAM50959.
XX
PT Producing para-hydroxy benzoic acid in green plant, comprises expression
PT of unique expression cassette containing gene encoding chorismate
PT pyruvate lyase operably linked to specific chloroplast targeting
PT sequence.
XX
PS Claim 17; Page 57; 60pp; English.
XX
CC The present nucleotide sequence is that of an open reading frame encoding
CC chloroplast-targeted chorismate pyruvate lyase (CPL) fusion protein TP-
CC CPL (see AAM50959). PCR (see ABA91841-42) was used to generate a DNA
CC fragment corresponding to the transit peptide of tomato Rubisco small
CC subunit and the first 4 amino acid residues of mature Rubisco. The DNA
CC fragment was ligated into pRT24a-CPL, which carries the Escherichia coli
CC CPL open reading frame (see ABA91837). The TP-CPL construct is an example
CC of expression cassettes of the invention that are designed for the high-
CC level production of p-hydroxybenzoic acid (pHBA) in higher plants.
CC Cleavage of TP-CPL in the chloroplast releases a novel polypeptide (see
CC AAM50961) that has full enzyme activity, converting chorismate to pHBA. A
CC claimed plant comprising a CPL expression cassette is selected from
CC soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat,
CC barley, oats, sorghum, rice, Arabidopsis, sugarcane, canola,
CC millet, bean, pea, rye, flax or a forage grass. pHBA is a monomeric
CC component of liquid crystal polymers which have application in the
CC automotive, electrical and other industries. (Updated on 29-AUG-2003 to
CC standardise OS field)
XX
SQ Sequence 684 BP; 160 A; 171 C; 179 G; 174 T; 0 U; 0 Other;

Query Match 100.0%; Score 684; DB 6; Length 684;
Best Local Similarity 100.0%; Pred. No. 1.4e-219;
Matches 684; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCTCTGTCTATTTCTTTCAGCAGCTGTGGCCACACGCGCAATGTTACACAAGCT 60
DB |||||
QY 61 AGCATGGTTGACCTTTCACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTTACAAAGAG 120
DB |||||
QY 61 AGCATGGTTGACCTTTCACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTTACAAAGAG 120
DB |||||
QY 121 CAAAACCTTGACATCACTTCCATCTCTAGCAATGTGGAAGAGTTAGTGTGATGCAAGTG 180
DB |||||
QY 121 CAAAACCTTGACATCACTTCCATCTCTAGCAATGTGGAAGAGTTAGTGTGATGCAAGTG 180
DB |||||
QY 181 TGGCATATGTACACCCCGCTTACGCAACTGCGTGGCTGCGCTATTGTAAAGAGATC 240
DB |||||
QY 181 TGGCATATGTACACCCCGCTTACGCAACTGCGTGGCTGCGCTATTGTAAAGAGATC 240
DB |||||
QY 241 CTGCGCTTGGATCCGCAACTGCTGACCTGGCTGTGCTGGAGATTCCATGACAAAAGCT 300
DB |||||
QY 241 CTGCGCTTGGATCCGCAACTGCTGACCTGGCTGTGCTGGAGATTCCATGACAAAAGCT 300
DB |||||
QY 301 TTTGAACAGCAGGAGAAAACGGTAAGCGTACGATGATCCGCGAAGGGTTTGTCTGAGCAG 360
DB |||||
QY 301 TTTGAACAGCAGGAGAAAACGGTAAGCGTACGATGATCCGCGAAGGGTTTGTCTGAGCAG 360
DB |||||
QY 361 AATGAATCCCGAAGAACTGCGCTGTCGCGAAGAGTCTCGTTACTGTTAGCTGAA 420
DB |||||
QY 361 AATGAATCCCGAAGAACTGCGCTGTCGCGAAGAGTCTCGTTACTGTTAGCTGAA 420
DB |||||
QY 421 ATTTTGTATGTGCGCATGGTGAACCGTGGCTGTGCGGTGCTGACCGTCTGCTCTGTGCA 480
DB |||||
QY 421 ATTTTGTATGTGCGCATGGTGAACCGTGGCTGTGCGGTGCTGACCGTCTGCTCTGTGCA 480
DB |||||
QY 481 AGTTTAAGCGGCGCGAGCTGCGTTACAAAATTTGGGTAAACCGCGTTAGGACGCTAT 540
DB |||||
QY 481 AGTTTAAGCGGCGCGAGCTGCGTTACAAAATTTGGGTAAACCGCGTTAGGACGCTAT 540
DB |||||
QY 541 CTGTTTCACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCCGTGATCCCGGGCTG 600
DB |||||

|||
541 CTGTTTCACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCCGTGATCCCGGGCTG 600
|||
601 TGGGGGCGACGTTCCCGCTCGATTAAAGCGGTAAACCGCTGTTCTTAACAGAACTGTTT 660
|||
601 TGGGGGCGACGTTCCCGCTCGATTAAAGCGGTAAACCGCTGTTCTTAACAGAACTGTTT 660
|||
661 TTACCGGCGTCACCGTTGTACTAA 694
|||
661 TTACCGGCGTCACCGTTGTACTAA 684
|||

RESULT 2
ABK47712
ID ABK47712 standard; DNA; 684 BP.
XX
AC ABK47712;
XX
DT 18-JUN-2002 (first entry)
XX
DE DNA sequence of ORF for TP-CPL fusion protein.
XX
KW SNG1; sinapoylglucose accumulator 1; SMT; aromatic ester; plant;
KW sinapoylglucose:malate sinapoyltransferase; glycosylated aromatic acid;
KW malate conjugated aromatic acid; polymer synthesis; ubiC; TP-CPL; gene;
KW carboxylic acid conjugated aromatic acid; chorismate pyruvate lyase;
KW tomato Rubisco small subunit precursor; rbcS2; mutant; ds.
XX
OS Escherichia coli.
OS Lycopersicon esculentum.
OS Synthetic.
OS Chimeric.
XX
FH Key Location/Qualifiers
FT CDS 1..684
FT /tag= a
FT /product= "TP-CPL fusion protein"
XX
PN WO200204653-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021283.
XX
PR 07-JUL-2000; 2000US-0216615P.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI Flint D, Meyer K, Viitanen PV;
XX
XX WPI; 2002-303779/34.
DR P-PSDB; AAU77942.
XX
XX Producing aromatic acid conjugates, involves contacting glycosylated
XX aromatic acid with malate, alpha-hydroxycarboxylic acid or alcohol in the
XX presence of sinapoylglucose:malate sinapoyltransferase.
XX
XX Example; Page 70; 72pp; English.
XX
CC The present invention relates to the isolation of Arabidopsis thaliana
CC gene (SNG1, sinapoylglucose accumulator 1) encoding
CC sinapoylglucose:malate sinapoyltransferase (SMT). SMT catalyses the
CC substitution of a glucose moiety on a glycosylated aromatic acid with a
CC malate moiety to form a malate conjugated aromatic acid. The enzyme is
CC useful for producing malate conjugated aromatic acids, carboxylic acid
CC conjugated aromatic acids or aromatic esters. The malate conjugated
CC aromatic acids are useful in the synthesis of various polymers. The
CC present sequence representing the ORF (open reading frame) for TP-CPL
CC (tomato Rubisco small subunit precursor for rbcS2-Escherichia coli
CC chorismate pyruvate lyase (CPL)) fusion protein is described in the
CC examples of the present invention
XX
SQ Sequence 684 BP; 160 A; 171 C; 179 G; 174 T; 0 U; 0 Other;
```

Query Match	100.0%;	Score 684;	DB 6;	Length 684;
Best Local Similarity	100.0%;	Pred. No. 1.4e-219;		
Matches 684;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	ATGCGTTCCTCTGTCATTTCTTCACGAGCTGTGGCCACACGCGAGCAATGTTACACAAGCT	60	
Db	1			
Qy	61	AGCATGGTTGCACTTTTCACTTGCTGTCTCAAAATCTTTCAGCCACATTTCCCTCTGTTACAAAAG	120	
Db	61			
Qy	121	CAAAACCTTTGACATCACTTCCATTTGCTGCAATGTTGGAAGAGTTAGCTGCAATGAGGTG	180	
Db	121			
Qy	181	TGSCATATGTACACACCCCGGTTTAAAGCAATCGCTGCGCTGCGCTATTTGTAAGAGATC	240	
Db	181			
Qy	241	CCTGCCCTTGGATCCGCAACTGCTCGACTGGCTGTTTCTTGGAGGATTTCCATGACAAAAGCT	300	
Db	241			
Qy	301	TTTGAACACGAGGGGAAAAACGGTTAAGCGGTGACGATGATCCCGCAAGGGTTTGTGAGCAG	360	
Db	301			
Qy	361	AATGAATCCCGAGAACTGCGGTGCTGCGCAAGAGTCTCGTTACTGGTTACGTGTA	420	
Db	361			
Qy	421	ATTTTGTATTGTGCCGATGGTGAAACCGTGGCTTTGCGCGTTCGTACCGTCTGTTCTGTGTGCA	480	
Db	421			
Qy	481	AGTTAAGCGGCGCGAGCTGCGTTTACAAAATTTGGGTAAACCGCGTTAGGACGCTAT	540	
Db	481			
Qy	541	CTGTTTCACATCATGCATTAACCCGGGACTTTATTGAGATAGGCCGTGATGCCGGGCTG	600	
Db	541			
Qy	601	TGGGGGCGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGTCTAAACAAGACTGTTT	660	
Db	601			
Qy	661	TTACCGGGCTGACCGTTGTACTAA 684		
Db	661			

RESULT 3	
ADC25988	
ID	ADC25988 standard; DNA; 684 BP.
XX	
XX	
AC	ADC25988;
XX	
AC	
DT	18-DEC-2003 (first entry)
XX	
DE	Tomato/ <i>Escherichia coli</i> chimeric TP-CPL DNA.
XX	
KW	UDP-glucosyltransferase; PHBA; p-hydroxybenzoic acid ester glucoside;
KW	liquid crystal polymer; LCP; methylparaben; preservative; food;
KW	cosmetic industry; ds; gene; tomato; TP-CPL; chimeric;
KW	chorismate pyruvate lyase; plant.
XX	
OS	Chimeric.
OS	Lycopersicon esculentum.
OS	<i>Escherichia coli</i> .
XX	
XX	

Key	Location/Qualifiers
FT CDS	1..684
FT	/*tag= a
FT	/product= "Tomato/Escherichia coli chimeric TP-CPL
FT	protein"
XX	
PN	WO2003066936-A2.
XX	
XX	14-AUG-2003.
XX	
PF	06-FEB-2003; 2003WO-US005863.
XX	
PR	07-FEB-2002; 2002US-035511P.
XX	
PA	(DUPO ) DU PONT DE NEMOURS & CO E I.
PI	Meyer K, Van Dyk DE, Viitanen PV;
XX	
XX	WPI; 2003-767259/72.
DR	P-PSDB; ADC25989.
XX	
PT	New nucleic acid encoding UDP-glucosyltransferase, useful for preparing
PT	cells that produce p-hydroxybenzoic acid glucose ester, also the new
PT	enzymes.
XX	
PS	Example 9; SEQ ID NO 41; 161bp; English.
XX	
CC	The invention relates to a novel isolated nucleic acid that encodes a UDP
CC	-glucosyltransferase. The method of the invention may be used to
CC	transform microorganisms or green plant cells so that these produce a
CC	higher level of high-volume chemicals or materials, such as pHBA (p-
CC	hydroxybenzoic acid) ester glucose which is an intermediate for pHBA, a
CC	monomer for liquid crystal polymers (LCPs) and starting material for
CC	methylparaben. Methylparaben is a preservative commonly used in the food
CC	and cosmetic industries. The encoded enzymes may be used for in vitro
CC	production of these compounds and for identifying similar enzymes by
CC	sequence comparison. The current sequence is that of the
CC	tomato/Escherichia coli chimeric TP-CPL (chorismate pyruvate lyase) DNA
CC	of the invention.
XX	
SQ	Sequence 684 BP; 160 A; 171 C; 179 G; 174 T; 0 U; 0 Other;
Query Match 100.0%; Score 684; DB 10; Length 684;	
Best Local Similarity 100.0%; Pred. No. 1.4e-219;	
Matches 684; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Oy	1 ATGGCTTCCTGTGCATTCTTCACGACGTGTTGCCACACGACGCAATGTTACACAAGCT 60
Db	1 ATGGCTTCCTGTGCATTCTTCACGACGTGTTGCCACACGACGCAATGTTACACAAGCT 60
Oy	61 AGCATGGTTGCACCTTTTCACGTGGTCTCAATCTTCAGCCACTTTCCTCTGTTACAAAGAG 120
Db	61 AGCATGGTTGCACCTTTTCACGTGGTCTCAATCTTCAGCCACTTTCCTCTGTTACAAAGAG 120
Oy	121 CAAACCTTGACATCACTTCCATTGTTAGCAATGGTGAAGAGTTAGTCGATCAGGTG 180
Db	121 CAAACCTTGACATCACTTCCATTGTTAGCAATGGTGAAGAGTTAGTCGATCAGGTG 180
Oy	181 TGGCATATGTCACACCCCGGTTHAACGCACTGCGTGGCTGCCGCTATTGTAAGAGATC 240
Db	181 TGGCATATGTCACACCCCGGTTHAACGCACTGCGTGGCTGCCGCTATTGTAAGAGATC 240
Oy	241 CTTGCCCTTGGATCCGAACTGCTCGACTGGCTGTTGTTGGAGGATTCATGACAAAACGT 300
Db	241 CTTGCCCTTGGATCCGAACTGCTCGACTGGCTGTTGTTGGAGGATTCATGACAAAACGT 300
Oy	301 TTTGAACAGCAGGGAAAAACGGTAAGCGTGACATGATCCGCGAAGGGTTGTGCGAGCAG 360
Db	301 TTTGAACAGCAGGGAAAAACGGTAAGCGTGACATGATCCGCGAAGGGTTGTGCGAGCAG 360
Oy	361 AATGAATCCCGAAGAACTCCGCGTCTGTCGCGAAGAGTCTCGTTACTGTTACGTGAA 420
Db	361 AATGAATCCCGAAGAACTCCGCGTCTGTCGCGAAGAGTCTCGTTACTGTTACGTGAA 420

421	Qy	ATTTGTTATGTGCGATGGTGAAACCGTGGCTTGC CGGTGCTACCGTGGTTCGTGTGCTCA	480
421	Db	ATTTGTTATGTGCGCATGGTGAAACCGTGGCTTGC CGGTGCTACCGTGGTTCGTGTGTCTCA	480
481	Qy	ACGTTAAGCGGCGCGAGCTGCGCTTACAAAATTTGGGTAAACCGCGTTAGAGACGCTAT	540
481	Db	ACGTTAAGCGGCGCGAGCTGCGCTTACAAAATTTGGGTAAACCGCGTTAGAGACGCTAT	540
541	Qy	CTGTTCACATCATCGACATTAACCGGACCTTTATTGAGATAGCGCGTGTATGCCGGGCTG	600
541	Db	CTGTTCACATCATCGACATTAACCGGACCTTTATTGAGATAGCGCGTGTATGCCGGGCTG	600
601	Qy	TGGGGGCGACGTTCCCGCTCGGATTAGCGGTAAACCGCTGTGTGCTTAAACAGAACTGTTTT	660
601	Db	TGGGGGCGACGTTCCCGCTCGGATTAGCGGTAAACCGCTGTGTGCTTAAACAGAACTGTTTT	660
661	Qy	TTACCGGCGTCAACGTTGTACTAA	684
661	Db	TTACCGGCGTCAACGTTGTACTAA	684

RESULT 4	
ADV91622	
ID	ADV91622 standard; DNA; 684 BP.
XX	
AC	
XX	ADV91622;
XX	
DT	10-MAR-2005 (first entry)
XX	
DE	Tomato rubisco small subunit precursor TP-E. coli CPL chimeric DNA.
XX	
KW	hydroxynone glucoaside; arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside;
KW	chormate pyruvate lyase; CPL; 4-hydroxycinnamoyl-CoA hydratase/lyase;
KW	HCHL; gene expression; hydroquinone; antiseptic; transgenic plant;
KW	genetically engineered macroorganism; antioxidant; antimicrobial agent;
KW	anti-inflammatory agent; carcinogenic inhibitor; melanoma; cytostatic;
KW	neoplasm; pHBA; para-hydroxybenzoic acid; p-hydroxybenzoic acid;
KW	rubisco small subunit precursor; transit peptide;
KW	chloroplast transit peptide; TP; chloroplast; gene; ds.
XX	
OS	Lycopersicon esculentum.
OS	Escherichia coli.
OS	Chimeric.

XX	US2004261147-A1.
PN	
XX	23-DEC-2004.
XX	
XX	16-JUN-2003; 2003US-00462162.
PF	
XX	16-JUN-2003; 2003US-00462162.
PR	
XX	(MEYE/) MEYER K.
XX	(VIIT/) VIITANEN P V.
PA	(FLIN/) FLINT D.
PA	
XX	Meyer K, Viitanen PV, Flint D;
PI	
XX	WPI; 2005-057232/06.
XX	P-PSDB; ADV91623.
XX	
XX	Producing hydroquinone glucoside in a green plant comprises growing a
PT	green plant having nucleic acid fragments and chorismate pyruvate lyase
PT	expression cassette.
PT	
XX	Example 4; SEQ ID NO 33; 70pp; English.
XX	
PS	

XX		The present invention relates to methods and materials to produce
CC	hydroquinone glucoside (arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside)	
CC	in genetically modified green plants and microorganisms. The method	
CC	relies upon transgenic plants or genetically modified microorganisms that	
CC	produce increased levels of the initial substrate para-hydroxybenzoic	
CC	acid (p-hydroxybenzoic acid; PHBA) in a biosynthetic pathway to produce	
CC	arbutin. Plants and microbes can be genetically engineered to produce	
CC	high levels of PHBA either by functional expression of the bacterial	
CC	protein chorismate pyruvate lyase (CPL) or by the expression of 4-	
CC	hydroxycinnamoyl-CoA hydratase/lyase (HCHL) or a combination of both.	
CC	Arbutin is useful as an antioxidant, antimicrobial agent, anti-	
CC	inflammatory agent and inhibitor of carcinogenesis (melanoma). The	
CC	present sequence is tomato rubisco small subunit precursor (zbcs2)	
CC	transit peptide (chloroplast transit peptide; TP)-Escherichia coli CPL	
CC	chimeric DNA. This sequence is present in the E. coli expression vector	
CC	construct pET24a-TP-CPL.	
XX		
SQ	Sequence 684 BP; 160 A; 171 C; 179 G; 174 T; 0 U; 0 Other;	
	Query Match	100.0%; Score 684; DB 14; Length 684;
	Best Local Similarity	100.0%; Pred. No. 1.4e-219;
	Matches 684; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
Qy	1 ATGCTTCCTCTGTCAATTTCTTACAGAGCTGTGGCCACACGCAGCAATGTTACACAAGCT 60	
Dd	1 ATGCTTCCTCTGTCAATTTCTTACAGAGCTGTGGCCACACGCAGCAATGTTACACAAGCT 60	
Qy	61 AGCATGTTGGACCTTTTCACTGTGTCCTCAAATCTTTCAGCCAATTTCCTGTTTACAAGAAG 120	
Dd	61 AGCATGTTGGACCTTTTCACTGTGTCCTCAAATCTTTCAGCCAATTTCCTGTTTACAAGAAG 120	
Qy	121 CAATAACCTTGACATCACTTCATTTGCTAGCAATGTTGAAGAGTTAGCTGCGATGACAGTG 180	
Dd	121 CAATAACCTTGACATCACTTCATTTGCTAGCAATGTTGAAGAGTTAGCTGCGATGACAGTG 180	
Qy	181 TGGCATATGTACACACCCCGGTTTAACGCCAACTGGCGTGCCTGCGCTATTGTTAAAGAGATC 240	
Dd	181 TGGCATATGTACACACCCCGGTTTAACGCCAACTGGCGTGCCTGCGCTATTGTTAAAGAGATC 240	
Qy	241 CTGCCCCCTGGATCCGGAACTGCTCGACTGGCTGTTGCTGGAGGATTCCATGACAAAACGT 300	
Dd	241 CTGCCCCCTGGATCCGGAACTGCTCGACTGGCTGTTGCTGGAGGATTCCATGACAAAACGT 300	
Qy	301 TTGTAAACAGCAGGGAAAAACGGTAAGCGTGCAGATGATCCCGAAGGGTTTTGTCGAGCAG 360	
Dd	301 TTGTAAACAGCAGGGAAAAACGGTAAGCGTGCAGATGATCCCGAAGGGTTTTGTCGAGCAG 360	
Qy	361 AATGAANAATCCCAGAAAGAACCTGCCCTGTGTCGCGAAGAGTCTCGTTACTGTTTACGTGAA 420	
Dd	361 AATGAANAATCCCAGAAAGAACCTGCCCTGTGTCGCGAAGAGTCTCGTTACTGTTTACGTGAA 420	
Qy	421 ATTTTGTTATGTGCCGATGTTGRAACCTGGCTTCCCGTCCGTACCGTCTTCTCTGTCTCA 480	
Dd	421 ATTTTGTATTGTGCCGATGTTGRAACCTGGCTTCCCGTCCGTACCGTCTTCTCTGTCTCA 480	
Qy	481 ACGTTAAGCGGGCCGGAGCTGGCGTTTACAAAAATTTGGGTAAAAACGCCGTTTAGGACGCTAT 540	
Dd	481 ACGTTAAGCGGGCCGGAGCTGGCGTTTACAAAAATTTGGGTAAAAACGCCGTTTAGGACGCTAT 540	
Qy	541 CTGTTTCAATCATCGACATTAACCCGGGACTTTTATTGAGATAGGCCGTGATGCCGGGCTG 600	
Dd	541 CTGTTTCAATCATCGACATTAACCCGGGACTTTTATTGAGATAGGCCGTGATGCCGGGCTG 600	
Qy	601 TGSGGGCGACGTTTCCCGCTCGCATTAAGCGGTAAACCGCTGTTGTGCTAACGAACTGTTT 660	
Dd	601 TGSGGGCGACGTTTCCCGCTCGCATTAAGCGGTAAACCGCTGTTGTGCTAACGAACTGTTT 660	
Qy	661 TTACCGCGCTCACCGTTGTACTAA 684	
Dd	661 TTACCGCGCTCACCGTTGTACTAA 684	



## RESULT 5

ADV91636  
 ID ADV91636 standard; DNA; 3452 BP.  
 AC  
 XX ADV91636;  
 XX  
 DT 10-MAR-2005 (first entry)  
 XX  
 DE C. parapsilosis pH8 1-H- E. coli CPL- A. thaliana UGT72B1 chimeric DNA.  
 XX  
 KW Hydroquinone glucoside; arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside;  
 KW chorismate pyruvate lyase; CPL; 4-hydroxycinnamoyl-CoA hydratase/lyase;  
 KW HCHL; gene expression; hydroquinone; antiseptic; transgenic plant;  
 KW genetically engineered microorganism; antioxidant; antimicrobial agent;  
 KW anti-inflammatory agent; carcinogenic inhibitor; melanoma; cytostatic;  
 KW neoplasm; PHBA 1-hydroxylase; pHBA 1-H; PHBA; para-hydroxybenzoic acid;  
 KW p-hydroxybenzoic acid; UDP-glucosyltransferase; UGT72B1; gene; db.  
 XX  
 OS Arabidopsis thaliana.  
 OS Candida parapsilosis.  
 OS Escherichia coli.  
 OS Chimeric.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 4..501  
 FT /\*tag= a  
 FT /product= "Escherichia coli CPL protein"  
 FT 541..1980  
 FT /\*tag= b  
 FT /product= "Candida parapsilosis pHBA 1-H protein"  
 FT /transl\_except= (pos: 559..561, aa:Trp)  
 FT CDS 2004..3446  
 FT /\*tag= c  
 FT /product= "Arabidopsis thaliana UGT72B1 protein"  
 XX  
 PN US2004261147-A1.  
 XX  
 XX 23-DEC-2004.  
 XX  
 XX 16-JUN-2003; 2003US-00462162.  
 XX  
 XX 16-JUN-2003; 2003US-00462162.  
 PR (MEYE/) MEYER K.  
 PA (VIIT/) VIITANEN P V.  
 PA (FLIN/) FLINT D.  
 XX  
 PI Meyer K, Viitanen PV, Flint D;  
 XX  
 XX WPI; 2005-057232/06.  
 DR P-PSDB; ADV91605, ADV91619, ADV91631.  
 XX  
 XX Producing hydroquinone glucoside in a green plant comprises growing a  
 PT green plant having nucleic acid fragments and chorismate pyruvate lyase  
 PT expression cassette.  
 XX  
 XX Example 7; SEQ ID NO 47; 70pp; English.  
 XX  
 CC The present invention relates to methods and materials to produce  
 CC hydroquinone glucoside (arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside)  
 CC in genetically modified green plants and microorganisms. The method  
 CC relies upon transgenic plants or genetically modified microorganisms that  
 CC produce increased levels of the initial substrate para-hydroxybenzoic  
 CC acid (p-hydroxybenzoic acid; PHBA) in a biosynthetic pathway to produce  
 CC arbutin. Plants and microbes can be genetically engineered to produce  
 CC high levels of PHBA either by functional expression of the bacterial  
 CC protein chorismate pyruvate lyase (CPL) or by the expression of 4-  
 CC hydroxycinnamoyl-CoA hydratase/lyase (HCHL) or a combination of both.  
 CC Arbutin is useful as an antioxidant, antimicrobial agent, anti-  
 CC inflammatory agent and inhibitor of carcinogenesis (melanoma). The  
 CC present sequence is candida parapsilosis pHBA 1-hydroxylase (pH8 1-H) -  
 CC Escherichia coli chorismate pyruvate lyase (CPL) and Arabidopsis thaliana  
 CC UDP-glucosyltransferase (UGT72B1) chimeric DNA.

XX

SQ Sequence 3452 BP; 932 A; 730 C; 856 G; 934 T; 0 U; 0 Other;

Query Match 73.2%; Score 501; DB 14; Length 3452;

Best Local Similarity 100.0%; Pred. No. 1.9e-157;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 CATATGTCACACCCCGCTTAACGCAACTGCGTGGCTGCGCTATTGTAAAGAGATCCCT 243

Db 1 CATATGTCACACCCCGCTTAACGCAACTGCGTGGCTGCGCTATTGTAAAGAGATCCCT 60

QY 244 GCCTGGATCCGCAACTGCTGACTGGCTGTTGCTGGAGGATTCATGACAAACGTTTT 303

Db 61 GCCTGGATCCGCAACTGCTGACTGGCTGTTGCTGGAGGATTCATGACAAACGTTTT 120

QY 304 GAACAGCAGGGGAAAAACGGTAAAGCTGACGATGATCCGCGAGGGTTTGTGAGCAGAAAT 363

Db 121 GAACAGCAGGGGAAAAACGGTAAAGCTGACGATGATCCGCGAGGGTTTGTGAGCAGAAAT 180

QY 364 GAAATCCCCGAAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAAT 423

Db 181 GAAATCCCCGAAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAAT 240

QY 424 TTGTTATGTCGCGATGGTGAACCGTGGCTTGGCGGTCGTACCGTCTGTTCTGTGTCAACG 483

Db 241 TTGTTATGTCGCGATGGTGAACCGTGGCTTGGCGGTCGTACCGTCTGTTCTGTGTCAACG 300

QY 484 TTAAGCGGCGCGAGCTGGCTTACAAAAATGGGTAAACCGCTAGGACGCTATCTG 543

Db 301 TTAAGCGGCGCGAGCTGGCTTACAAAAATGGGTAAACCGCTAGGACGCTATCTG 360

QY 544 TTCAATCATCGACATTAAACCGGAGCTTTATTAGATAGGCGGTGATGCCGGCTGTGG 603

Db 361 TTCAATCATCGACATTAAACCGGAGCTTTATTAGATAGGCGGTGATGCCGGCTGTGG 420

QY 604 GGGCGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTA 663

Db 421 GGGCGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTA 480

QY 664 CCGGCGTCACCGTTGTTACTAA 684

Db 481 CCGGCGTCACCGTTGTTACTAA 501

## RESULT 6

ADV91643

ID ADV91643 standard; DNA; 1971 BP.

AC ADV91643;

XX

XX 10-MAR-2005 (first entry)

DT

XX C. parapsilosis pH8 1-H- E. coli CPL chimeric DNA.

XX Hydroquinone glucoside; arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside;

XX chorismate pyruvate lyase; CPL; 4-hydroxycinnamoyl-CoA hydratase/lyase;

XX HCHL; gene expression; hydroquinone; antiseptic; transgenic plant;

XX genetically engineered microorganism; antioxidant; antimicrobial agent;

XX anti-inflammatory agent; carcinogenic inhibitor; melanoma; cytostatic;

XX neoplasm; PHBA 1-hydroxylase; pHBA; para-hydroxybenzoic acid;

XX p-hydroxybenzoic acid; gene; db.

OS Candida parapsilosis.

OS Escherichia coli.

OS Chimeric.

XX Key Location/Qualifiers

FT CDS 4..501

FT /\*tag= a

FT /product= "Escherichia coli CPL protein"

FT /transl\_except= (pos: 46..48, aa:Cys)

FT CDS 522..1961

FT /\*tag= b

FT /product= "Candida parapsilosis pHBA 1-H protein"  
FT /transl\_except= (pos: 540..542, aa:Trp)  
PN US2004261147-A1.  
XX 23-DEC-2004.  
XX 16-JUN-2003; 2003US-00462162.  
XX 16-JUN-2003; 2003US-00462162.  
XX (MEYE/) MEYER K.  
PA (VIIT/) VIITANEN P V.  
PA (FLIN/) FLINT D.  
XX Meyer K, Viitanen PV, Flint D;  
XX WPI; 2005-057232/06.  
DR P-PSDB; ADV91605, ADV91619.  
XX Producing hydroquinone glucoside in a green plant comprises growing a  
PT green plant having nucleic acid fragments and chorismate pyruvate lyase  
PT expression cassette.  
XX Example 7; SEQ ID NO 54; 70pp; English.  
XX The present invention relates to methods and materials to produce  
CC hydroquinone glucoside (arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside)  
CC in genetically modified green plants and microorganisms. The method  
CC relies upon transgenic plants or genetically modified microorganisms that  
CC produce increased levels of the initial substrate para-hydroxybenzoic  
CC acid (p-hydroxybenzoic acid; pHBA) in a biosynthetic pathway to produce  
CC arbutin. Plants and microbes can be genetically engineered to produce  
CC high levels of pHBA either by functional expression of the bacterial  
CC protein chorismate pyruvate lyase (CPL) or by the expression of 4-  
CC hydroxycinnamoyl-CoA hydratase/lyase (HCHL) or a combination of both.  
CC Arbutin is useful as an antioxidant, antimicrobial agent, anti-  
CC inflammatory agent and inhibitor of carcinogenesis (melanoma). The  
CC present sequence is Candida parapsilosis pHBA 1-hydroxylase (pHB 1-H) and  
CC Escherichia coli chorismate pyruvate lyase (CPL) chimeric DNA.  
XX  
SQ Sequence 1971 BP; 550 A; 383 C; 484 G; 554 T; 0 U; 0 Other;  
  
Query Match 73.0%; Score 499.4; DB 14; Length 1971;  
Best Local Similarity 99.8%; Pred. No. 4.9e-157;  
Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 184 CATATGTACACCCCGGTTAAGCAACTGCGTGGCTGCGCTATTGTAAAGAGATCCCT 243  
DB 1 CATATGTACACCCCGGTTAAGCAACTGCGTGGCTGCGCTATTGTAAAGAGATCCCT 60  
  
QY 244 GCCCTGGATCCGAACCTGCTGACCTGGCTGTGGAGGATTCATGACAAAAGCTTTT 303  
DB 61 GCCCTGGATCCGAACCTGCTGACCTGGCTGTGGAGGATTCATGACAAAAGCTTTT 120  
  
QY 304 GAACAGCAGGAAAAACGGTAAGCTGACGATGATCCGGAAGGTTTGTGAGCAGAAAT 363  
DB 121 GAACAGCAGGAAAAACGGTAAGCTGACGATGATCCGGAAGGTTTGTGAGCAGAAAT 180  
  
QY 364 GAAATCCCGAAGAACTGCGCTGCTGCGGAAAGAGTCTCGTTACTGGTTACGTGAATTT 423  
DB 181 GAAATCCCGAAGAACTGCGCTGCTGCGGAAAGAGTCTCGTTACTGGTTACGTGAATTT 240  
  
QY 424 TTGTTATGTGCGGATGGTGAACCGTGGCTGTGCGGTCGTACCGTGGTCTGTGTCAACG 483  
DB 241 TTGTTATGTGCGGATGGTGAACCGTGGCTGTGCGGTCGTACCGTGGTCTGTGTCAACG 300  
  
QY 484 TTAACGGCGCGGAGCTGCGCTTACAAAATTTGGGTAAACGCCCTTAGGACGCTATCTG 543  
DB 301 TTAACGGCGCGGAGCTGCGCTTACAAAATTTGGGTAAACGCCCTTAGGACGCTATCTG 360  
  
QY 544 TTCACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCCGTGATGCCGGCTGTGG 603  
|||||

DB 361 TTCACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCCGTGATGCCGGCTGTGG 420  
QY 604 GGGGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTTA 663  
XX |||||||  
DB 421 GGGGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTTA 480  
XX |||||||  
QY 664 CCGGCGTCACCGTTGTACTAA 684  
XX |||||||  
DB 481 CCGGCGTCACCGTTGTACTAA 501  
XX |||||||  
  
RESULT 7  
ADV91618  
ID ADV91618 standard; DNA; 498 BP.  
XX AC  
XX ADV91618;  
XX 10-MAR-2005 (first entry)  
XX Escherichia coli chorismate pyruvate lyase (CPL) DNA.  
DE  
XX Hydroquinone glucoside; arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside;  
KW chorismate pyruvate lyase; CPL; 4-hydroxycinnamoyl-CoA hydratase/lyase;  
KW HCHL; gene expression; hydroquinone; antioxidant; transgenic plant;  
KW genetically engineered microorganism; antioxidant; antimicrobial agent;  
KW anti-inflammatory agent; carcinogenic inhibitor; melanoma; cytostatic;  
KW neoplasm; pHBA; para-hydroxybenzoic acid; p-hydroxybenzoic acid; gene;  
KW ds.  
XX Escherichia coli.  
FH Key Location/Qualifiers  
FT CDS 1..498  
FT /\*tag= a  
FT /product= "Escherichia coli CPL protein"  
XX  
XX US2004261147-A1.  
XX 23-DEC-2004.  
XX 16-JUN-2003; 2003US-00462162.  
XX 16-JUN-2003; 2003US-00462162.  
XX (MEYE/) MEYER K.  
PA (VIIT/) VIITANEN P V.  
PA (FLIN/) FLINT D.  
XX Meyer K, Viitanen PV, Flint D;  
XX WPI; 2005-057232/06.  
DR P-PSDB; ADV91619.  
XX  
XX Producing hydroquinone glucoside in a green plant comprises growing a  
PT green plant having nucleic acid fragments and chorismate pyruvate lyase  
PT expression cassette.  
XX Example 4; SEQ ID NO 29; 70pp; English.  
XX The present invention relates to methods and materials to produce  
CC hydroquinone glucoside (arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside)  
CC in genetically modified green plants and microorganisms. The method  
CC relies upon transgenic plants or genetically modified microorganisms that  
CC produce increased levels of the initial substrate para-hydroxybenzoic  
CC acid (p-hydroxybenzoic acid; pHBA) in a biosynthetic pathway to produce  
CC arbutin. Plants and microbes can be genetically engineered to produce  
CC high levels of pHBA either by functional expression of the bacterial  
CC protein chorismate pyruvate lyase (CPL) or by the expression of 4-  
CC hydroxycinnamoyl-CoA hydratase/lyase (HCHL) or a combination of both.  
CC Arbutin is useful as an antioxidant, antimicrobial agent, anti-  
CC inflammatory agent and inhibitor of carcinogenesis (melanoma). The  
CC present sequence is Escherichia coli chorismate pyruvate lyase (CPL) DNA.  
CC This sequence is found in the E. coli expression vector construct

CC pET24a(+).

XX Sequence 498 BP; 114 A; 122 C; 141 G; 121 T; 0 U; 0 Other;

XX

Query Match 72.8%; Score 498; DB 14; Length 498;  
Best Local Similarity 100.0%; Pred. No. 7.1e-157; Indels 0; Gaps 0;  
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 ATGTCAACCCCCGTTAAAGCAACTCGTGCCTGCTATTTGAAGAAGATCCCTGCC 246  
DB 1 ATGTCAACCCCCGTTAAAGCAACTCGTGCCTGCTATTTGAAGAAGATCCCTGCC 60

QY 247 CTGGATCGGCACTGCTCGACTGCTGTCGTGAGAGATTCATGACAAAACGTTTTGAA 306  
DB 61 CTGGATCGGCACTGCTCGACTGCTGTCGTGAGAGATTCATGACAAAACGTTTTGAA 120

QY 307 CAGCAGGAAAAACGGTAAGCGTGACGATGATCCGCAAGGTTTGTGCGAGCAGAATGAA 366  
DB 121 CAGCAGGAAAAACGGTAAGCGTGACGATGATCCGCAAGGTTTGTGCGAGCAGAATGAA 180

QY 367 ATCCCCGAAGAACCTCGCGCTGCTCCGAAAGAGTCTGTTACTGTTACGTGAAATTTTG 426  
DB 181 ATCCCCGAAGAACCTCGCGCTGCTCCGAAAGAGTCTGTTACTGTTACGTGAAATTTTG 240

QY 427 TTATGTCCGAGTGAACCGTGGCTTGCCTGCGGTGATCCGTCGTCTGTGTCACAGTTA 486  
DB 241 TTATGTCCGAGTGAACCGTGGCTTGCCTGCGGTGATCCGTCGTCTGTGTCACAGTTA 300

QY 487 AGCGGCGCGAGCTGGCTTACAAAATTGGTTAAACCGCGTTAGCACCTATCTGTTTC 546  
DB 301 AGCGGCGCGAGCTGGCTTACAAAATTGGTTAAACCGCGTTAGCACCTATCTGTTTC 360

QY 547 ACATCATCGACATTAACCCGGACTTTATTTAGATAGCGCTGATGCCGGCTGTGGGG 606  
DB 361 ACATCATCGACATTAACCCGGACTTTATTTAGATAGCGCTGATGCCGGCTGTGGGG 420

QY 607 CGAGCTTCCCCTCGCGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTACCG 666  
DB 421 CGAGCTTCCCCTCGCGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTACCG 480

QY 667 GCGTCACCGTTGACTAA 584  
DB 481 GCGTCACCGTTGACTAA 498

RESULT 8  
AAD06956  
ID AAD06956 standard; DNA; 854 BP.  
XX AC  
XX AAD06956;  
XX  
DT 06-AUG-2001 (first entry)  
XX DE  
XX ubiC gene encoding chorismate pyruvate lyase.

XX p-hydroxybenzoic acid; pHb; aromatic pathway; carbon source; ubiC;  
KW chorismate pyruvate lyase; psrA; phosphoenolpyruvate synthase;  
KW PEP synthase; tktA; transketolase; aroC; chorismate synthase; aroA;  
KW 5-enolpyruvoylshikimate-3-phosphate synthase; EpsP synthase; aroI;  
KW shikimate kinase II; aroB; 3-dehydroquininate synthase; DHQ synthase; aroG;  
XX 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase; DHAP synthase; ds.  
OS Escherichia coli.  
XX  
XX US6210937-B1.  
PN  
XX 03-APR-2001.  
PD  
XX 22-APR-1998; 98US-00064693.  
PF  
XX 22-APR-1997; 97US-0044094P.  
PR  
XX (BECH-) BECHTEL BWXT IDAHO LLC.  
PA

ID	AA06957 standard; DNA; 6641 BP.	Qy	367	ATCCCCGAGAACTGCGCTCTGCGGAAAGAGTCTGTTACTGTTACGTTAGTGAATTTTG	426
XX	AA06957;	Db	501	ATCCCCGAGAACTGCGCTCTGCGGAAAGAGTCTGTTACTGTTACGTTAGTGAATTTTG	560
AC					
DT	06-AUG-2001 (first entry)	Qy	427	TTATGTGCCGATGTTGAACCGTGGCTTGCCTGCTACCGTCTGTTCCCTGTTGTCAACGTTA	486
DE	pME2 comprising ubiC, aroG, tkTA and ppsA genes of the aromatic pathway.	Db	561	TTATGTGCCGATGTTGAACCGTGGCTTGCCTGCTACCGTCTGTTCCCTGTTGTCAACGTTA	620
XX					
KW	p-hydroxybenzoic acid; pHb; aromatic pathway; carbon source; ubiC;	Qy	487	AGCGGGCGGAGCTGGCGTTACAAAAATTGGTAAAAACCGCGTTAGGACGCTATCTGTTTC	546
KW	chorismate pyruvate lyase; ppsA; phosphoenolpyruvate synthase;	KW	621	AGCGGGCGGAGCTGGCGTTACAAAAATTGGTAAAAACCGCGTTAGGACGCTATCTGTTTC	680
KW	PEP synthase; tkTA; transketolase; aroC; chorismate synthase; aroA;	Db			
KW	5-enolpyruvylshikimate-3-phosphate synthase; BPS synthase; aroL;	Qy	547	ACATCATCGACATTAACCGGAGCTTTATTGAGATAGCCCGCTGATGCCGCGCTGTGGGG	606
KW	shikimate kinase II; aroB; 3-dehydroquinate synthase; DHQ synthase; aroG;	Db	681	ACATCATCGACATTAACCGGAGCTTTATTGAGATAGCCCGCTGATGCCGCGCTGTGGGG	740
KW	3-deoxy-D-arabino-heptulosonate-7-phosphate synthase; DHAP synthase;				
XX	pME2; db.				
OS	Escherichia coli.	Qy	607	CGAGTTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTTACCG	666
XX		Db	741	CGAGTTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTTACCG	800
PN	US6210937-B1.				
PD	03-APR-2001.				
XX		Qy	667	CGTCACCGTTGTACTAA	684
XX		Db	801	CGTCACCGTTGTACTAA	818
PF	22-APR-1998; 98US-00064693.				
XX					
PR	22-APR-1997; 97US-0044094P.				
XX					
PA	(BECH-) BECHTEL BWXT IDAHO LLC.				
XX					
PI	Ward TE, Watkins CS, Bulmer DK, Johnson BF, Amaratunga M;				
XX					
DR	WPI; 2001-280857/29.				
XX		DT	17-JAN-1997 (first entry)		
XX		XX			
PT	Producing aromatic compounds, especially commercially acceptable levels	DE			
PT	of p-hydroxybenzoic acid, comprises developing genetically engineered	XX			
PT	bacteria that carry selected genes of the common aromatic pathway.	XX			
XX		KW			
PS	Claim 4; Col 31-40; 25pp; English.	KW			
XX		KW			
CC	The present invention relates to a method of producing p-hydroxybenzoic	OS			
CC	acid (pHB) which comprises transforming microorganisms with plasmids	XX			
CC	carrying selected genes of the common aromatic pathway and by conversion	XX			
CC	of carbon sources. The selected genes of the common aromatic pathway of	XX			
CC	the invention includes ubiC gene encoding chorismate pyruvate lyase, aroG	XX			
CC	structural gene encoding 3-deoxy-D-arabino-heptulosonate-7-phosphate	XX			
CC	(DHAP) isoenzyme synthase (phe), tkTA gene encoding transketolase, ppsA	XX			
CC	gene encoding phosphoenolpyruvate (PEP) synthase, aroC gene encoding	XX			
CC	chorismate synthase, aroL gene encoding shikimate kinase II, aroA gene	XX			
CC	encoding 5-enolpyruvylshikimate-3-phosphate synthase and aroB	XX			
CC	gene encoding 3-dehydroquinate (DHQ) synthase. The method is used for	XX			
CC	converting chorismate to a selected aromatic compound. The method is	XX			
CC	particularly used for the production of commercially acceptable levels of	XX			
CC	aromatic compounds, particularly p-hydroxybenzoic acid. The present	XX			
CC	invention is pME2 plasmid which comprises ubiC, aroG, tkTA and ppsA genes	XX			
CC	involved in the aromatic pathway	XX			
XX		XX			
SQ	Sequence 6641 BP; 1567 A; 1698 C; 1853 G; 1523 T; 0 U; 0 Other;	XX			
		PN	JP08107789-A.		
		XX			
		XX	30-APR-1996.		
		XX			
		XX	13-OCT-1994; 94JP-00273071.		
		XX			
		PR	13-OCT-1994; 94JP-00273071.		
		XX			
		PA	(ALPH-) ALPHA SHOKUHIN KK.		
		XX			
		DR	WPI; 1996-262599/27.		
		DR	P-PSDB; AAR97746, AAR97747.		
		XX			
		PT	Genes coding for ubiquinone biosynthetic enzymes - useful for prodn. of		
		PT	ubiquinone-10 by transformed photosynthetic bacteria.		
		XX			
		PS	Claim 1; Fig 2; 11pp; Japanese.		

Qy	187	ATGTCACACCCCGCTTACCGCACTGCGTGGCTGCGCTATTGTTAAAGAGATCCCTGCC	246
Db	321	ATGTCACACCCCGCTTACCGCACTGCGTGGCTGCGCTATTGTTAAAGAGATCCCTGCC	380
Qy	247	CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAAAGTTTGAA	306
Db	381	CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAAAGTTTGAA	440
Qy	307	CAGCAGGGAACAAACGGTAAGCGTACGATGATCCCGAAGGTTTGTTCGACAGATGAA	366
Db	441	CAGCAGGGAACAAACGGTAAGCGTACGATGATCCCGAAGGTTTGTTCGACAGATGAA	500

Query Match 72.8%; Score 498; DB 4; Length 6641;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-156;  
 Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX The present sequence comprises the ubiquinone biosynthetic enzyme genes  
 CC ubiC and ubiA, which were cloned from the chromosomal DNA of *E. coli*  
 CC (Kohara map phage DNA bank IP8 (634); Cell 50, 495-508 (1987)). A large  
 CC amt. of ubiquinone-10 can be produced by culturing photosynthetic  
 CC bacteria transformed with the novel plasmid pRSPAC, which contains the  
 CC ubiC and ubiA genes and the glutamate synthase gene promoter,  
 CC specifically the Rhodobacter capsulatus transformant MC9R/pRSPAC  
 XX  
 SQ Sequence 2000 BP; 444 A; 462 C; 538 G; 556 T; 0 U; 0 Other;

Query Match 72.6%; Score 496.4; DB 2; Length 2000;  
 Best Local Similarity 99.8%; Pred. No. 5.1e-156;  
 Matches 497; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 187 ATGTACACCCCGCTTAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATCCCTGCC 246  
 Db 380 ATGTACACCCCGCTTAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATCCCTGCC 439  
 QY 247 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAACGTTTGA 306  
 Db 440 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAACGTTTGA 499  
 QY 307 CAGCAGGGAACCGTAAAGCGTACGATATCCGGAAGGGTTTTCGAGCAGAGATGAA 366  
 Db 500 CAGCAGGGAACCGTAAAGCGTACGATATCCGGAAGGGTTTTCGAGCAGAGATGAA 559  
 QY 367 ATCCCGGAAGACTGCGCTGCTGCCGAAGAGTCTCGTTACTGTTAGTGAATTTG 426  
 Db 560 ATCCCGGAAGACTGCGCTGCTGCCGAAGAGTCTCGTTACTGTTAGTGAATTTG 619  
 QY 427 TTATGTCCGATGGTGAACCGTGGCTTGGCGGTGCTACCGTCTGTTCTGTCAACGTTA 486  
 Db 620 TTATGTCCGATGGTGAACCGTGGCTTGGCGGTGCTACCGTCTGTTCTGTCAACGTTA 679  
 QY 487 AGCGGCGGAGTGGGTTACAAAATTGGGTAAACCGGTTAGGACGCTATCTGTTTC 546  
 Db 680 AGCGGCGGAGTGGGTTACAAAATTGGGTAAACCGGTTAGGACGCTATCTGTTTC 739  
 QY 547 ACATCATCAGATTAACCGGACTTTATTGAGTAGGCGGTGATCCCGGGCTGTGGGG 606  
 Db 740 ACATCATCAGATTAACCGGACTTTATTGAGTAGGCGGTGATCCCGGGCTGTGGGG 799  
 QY 607 CGACGTTCCCGCTCGGATTAAAGCGGTAAACCGCTGTGTTACAGAACTGTTTTACCG 666  
 Db 800 CGACGTTCCCGCTCGGATTAAAGCGGTAAACCGCTGTGTTACAGAACTGTTTTACCG 859  
 QY 667 GCGTACCGTTGTTACTAA 684  
 Db 860 GCGTACCGTTGTTACTAA 877

RESULT 11  
 AAQ92409  
 ID AAQ92409 standard; DNA; 495 BP.  
 XX  
 AC AAQ92409;  
 XX  
 DT 11-JAN-1996 (first entry)  
 XX  
 DE *E. coli* chorismate-pyruvate lyase coding sequence.  
 XX  
 KW Chorismate-pyruvate lyase gene; ubiC gene; transgenic plant;  
 KW tobacco mosaic virus resistance; antiviral activity;  
 KW p-hydroxybenzoic acid; B8.  
 XX  
 OS *Escherichia coli*.  
 XX  
 PN DE4423022-Cl.  
 XX  
 PD 24-MAY-1995.  
 XX  
 PF 30-JUN-1994; 94DE-04423022.

XX 30-JUN-1994; 94DE-04423022.  
 XX (HEID/) HEIDE L.  
 XX Heide L, Siebert M, Severin K;  
 XX WPI; 1995-186908/25.  
 XX P-PSDB; AAR74742.  
 XX Transgenic plants with increased content of resistance factor - contg.  
 PT bacterial DNA coding for enzyme catalysing prodn. of resistance factor,  
 PT e.g. antiviral p-hydroxy-benzoic acid.  
 XX  
 PS Claim 3; Page 5; 7pp; German.  
 XX  
 CC Transgenic plants containing a bacterial gene which encodes an enzyme  
 CC able to catalyse prodn. of an antiviral, bactericidal, fungicidal or  
 CC insecticidal factor are new. A preferred gene is ubiC from *E. coli*  
 CC (AAQ92409) which codes for chorismate-pyruvate lyase (AAR74742). The  
 CC lyase catalyses conversion of chorismate to p-hydroxybenzoic acid and  
 CC transgenic tobacco plants which express the active enzyme are resistant  
 CC to tobacco mosaic virus  
 XX  
 SQ Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;  
 Query Match 72.4%; Score 495; DB 2; Length 495;  
 Best Local Similarity 100.0%; Pred. No. 7.2e-156;  
 Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 187 ATGTACACCCCGCTTAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATCCCTGCC 246  
 Db 1 ATGTACACCCCGCTTAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATCCCTGCC 60  
 QY 247 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAACGTTTGA 306  
 Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAACGTTTGA 120  
 QY 307 CAGCAGGGAACCGTAAAGCGTACGATATCCGGAAGGGTTTTCGAGCAGAGATGAA 366  
 Db 121 CAGCAGGGAACCGTAAAGCGTACGATATCCGGAAGGGTTTTCGAGCAGAGATGAA 180  
 QY 367 ATCCCGGAAGACTGCGCTGCTGCCGAAGAGTCTCGTTACTGTTAGTGAATTTG 426  
 Db 181 ATCCCGGAAGACTGCGCTGCTGCCGAAGAGTCTCGTTACTGTTAGTGAATTTG 240  
 QY 427 TTATGTCCGATGGTGAACCGTGGCTGCTACCGGTGCTGTTCTGTGTTCAACGTTA 486  
 Db 241 TTATGTCCGATGGTGAACCGTGGCTGCTGTTCTGTGTTCAACGTTA 300  
 QY 487 AGCGGCGGAGTGGCGTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 546  
 Db 301 AGCGGCGGAGTGGCGTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 360  
 QY 547 ACATCATCAGATTAACCGGACTTTATTGAGTAGGCGGTGATCCCGGGCTGTGGGG 606  
 Db 361 ACATCATCAGATTAACCGGACTTTATTGAGTAGGCGGTGATCCCGGGCTGTGGGG 420  
 QY 607 CGACGTTCCCGCTCGGATTAAAGCGGTAAACCGCTGTGTTACAGAACTGTTTTACCG 666  
 Db 421 CGACGTTCCCGCTCGGATTAAAGCGGTAAACCGCTGTGTTACAGAACTGTTTTACCG 480  
 QY 667 GCGTACCGTTGTTACTAC 681  
 Db 481 GCGTACCGTTGTTACTAC 495  
 RESULT 12  
 ABA91837  
 ID ABA91837 standard; DNA; 495 BP.  
 XX  
 AC ABA91837;  
 XX

DT 15-MAY-2002 (first entry)  
 XX Escherichia coli chorismate pyruvate lyase gene coding region.  
 DE  
 XX Chorismate pyruvate lyase; CPL; enzyme; p-hydroxybenzoic acid;  
 KW transgenic plant; ubiC; gene; ds.  
 KW  
 XX Escherichia coli.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..495  
 FT /\*tag= a  
 FT /partial  
 FT /product= "chorismate\_pyruvate\_lyase"  
 FT /note= "the CDS does not include a stop codon"  
 XX  
 XX WO200194607-A2.  
 XX  
 XX 13-DEC-2001.  
 XX  
 XX 22-MAY-2001; 2001WO-US016661.  
 XX  
 XX 02-JUN-2000; 2000US-0209854P.  
 XX  
 XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX  
 XX Meyer K, Van Dyk DE, Viitanen PV;  
 PI  
 XX WPI; 2002-226795/28.  
 DR P-PSDB; AAM50958.  
 DR  
 XX Producing para-hydroxy benzoic acid in green plant, comprises expression  
 PT of unique expression cassette containing gene encoding chorismate  
 PT pyruvate lyase operably linked to specific chloroplast targeting  
 PT sequence.  
 XX  
 XX Claim 4; Page 55; 60pp; English.  
 PS  
 XX The present sequence is that of the coding region of the Escherichia coli  
 CC strain W3110 ubiC gene in expression construct pET24a. The gene encodes  
 CC chorismate pyruvate lyase (CPL, see AAM50968). It was obtained by PCR  
 CC amplification of strain W3110 genomic DNA using primers (see ABA91839-40)  
 CC based on the published E. coli ubiC gene. The CPL open reading frame is  
 CC used in an expression cassette designed for the high-level production of  
 CC p-hydroxybenzoic acid (pHBA) in green plants. The expression cassette  
 CC comprises the CPL coding sequence operably linked to a promoter capable  
 CC of driving protein expression in higher plants. The cassette also has a  
 CC sequence encoding a chloroplast transit peptide, its natural cleavage  
 CC site, and a small portion of a transit peptide donor protein fused to the  
 CC N-terminus of CPL. The chloroplast targeting sequence targets the foreign  
 CC protein to the chloroplast and aids in its uptake into the organelle. The  
 CC cleavage site is unique to the transit peptide, and cleavage of the  
 CC chimeric protein encoded by the cassette at this site releases a novel  
 CC polypeptide that has full enzyme activity, comprising the mature CPL  
 CC enzyme and a small portion of transit peptide donor (see AAM50961). A  
 CC plant comprising the CPL expression cassette is claimed, and may be  
 CC soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat,  
 CC barley, oats, sorghum, rice, Arabidopsis, sugarbeet, sugarcane, canola,  
 CC millet, bean, pea, rye, flax or a forage grass. pHBA is a monomeric  
 CC component of liquid crystal polymers which have application in the  
 CC automotive, electrical and other industries  
 XX  
 SQ Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;  
 Query Match 72.4%; Score 495; DB 6; Length 495;  
 Best Local Similarity 100.0%; Pred. No. 7.2e-156;  
 Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 187 ATGTACACCCCGCGTTAAACGCACTGGTGGCTGCGCTATTGTAAAGAGATCCCTGCC 246  
 DB 1 ATGTACACCCCGCGTTAAACGCACTGGTGGCTGCGCTATTGTAAAGAGATCCCTGCC 60  
 OY 247 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCCATGACAAACGTTTGA 306

Db 61 CTGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAACGTTTGA 120  
 OY 307 CAGCAGGGAACGCGTAAGCGTGACGATGATCCGCGAAGGTTTGTGCGAGCAGATGAA 366  
 Db 121 CAGCAGGGAACGCGTAAGCGTGACGATGATCCGCGAAGGTTTGTGCGAGCAGATGAA 180  
 OY 367 ATCCCGGAAGAACTGCGCGTCTGCTCCGAAAGAGTCTGTTACTGTTACGTGAAATTTTG 426  
 Db 181 ATCCCGGAAGAACTGCGCGTCTGCTCCGAAAGAGTCTGTTACTGTTACGTGAAATTTTG 240  
 OY 427 TTATGTCCGATGTTGAACCGTGCTGCGCGTCTGCTCCGAAAGAGTCTGTTACTGTTACGTGAAATTTTG 486  
 Db 241 TTATGTCCGATGTTGAACCGTGCTGCGCGTCTGCTCCGAAAGAGTCTGTTACTGTTACGTGAAATTTTG 300  
 OY 487 AGCGGGCGGAGCTGGCGTTACAAAATTGGGTTAAACCGCGTTAGGACGCTATCTGTTTC 546  
 Db 301 AGCGGGCGGAGCTGGCGTTACAAAATTGGGTTAAACCGCGTTAGGACGCTATCTGTTTC 360  
 OY 547 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGCCCGTGTATGCCGGCTGTGGGG 606  
 Db 361 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGCCCGTGTATGCCGGCTGTGGGG 420  
 OY 607 CGAGTTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTTACCG 666  
 Db 421 CGAGTTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTTACCG 480  
 OY 667 GCGTCACCGTTGTATAC 681  
 Db 481 GCGTCACCGTTGTATAC 495

RESULT 13  
 ABK47709  
 ID ABK47709 standard; DNA; 495 BP.  
 XX  
 AC ABK47709;  
 XX  
 DT 18-JUN-2002 (first entry)  
 XX  
 DE DNA sequence of ORF for CPL in pET24a E. coli expression construct.  
 XX  
 KW SNGI; sinapoylglucose accumulator 1; SMT; aromatic ester;  
 KW sinapoylglucose:malate sinapoyltransferase; glycosylated aromatic acid;  
 KW malate conjugated aromatic acid; polymer synthesis; ubiC; CPL; gene;  
 KW carboxylic acid conjugated aromatic acid; chorismate pyruvate lyase;  
 KW pET24a; ds.  
 XX  
 OS Escherichia coli.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..495  
 FT /\*tag= a  
 FT /partial  
 FT /product= "CPL"  
 FT /note= "This sequence lacks a stop codon"  
 XX  
 XX WO200204653-A2.  
 XX  
 XX 17-JAN-2002.  
 XX  
 XX 05-JUL-2001; 2001WO-US021283.  
 XX  
 XX 07-JUL-2000; 2000US-0216615P.  
 XX  
 XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX  
 XX Flint D, Meyer K, Viitanen PV;  
 XX  
 XX WPI; 2002-303779/34.  
 DR P-PSDB; AAU77941.  
 XX  
 XX Producing aromatic acid conjugates, involves contacting glycosylated

PT aromatic acid with malate, alpha-hydroxycarboxylic acid or alcohol in the  
 PT presence of sinapoylglucose:malate sinapoyltransferase.

XX  
 PS  
 XX  
 XX Example; Page 68; 72pp; English.

CC The present invention relates to the isolation of Arabidopsis thaliana  
 CC gene (SNGL, sinapoylglucose accumulator 1) encoding  
 CC sinapoylglucose:malate sinapoyltransferase (SMT). SMT catalyses the  
 CC substitution of a glucose moiety on a glycosylated aromatic acid with a  
 CC malate moiety to form a malate conjugated aromatic acid. The enzyme is  
 CC useful for producing malate conjugated aromatic acids, carboxylic acid  
 CC conjugated aromatic acids or aromatic esters. The malate conjugated  
 CC aromatic acids are useful in the synthesis of various polymers. The  
 CC present sequence representing the ORF (open reading frame) for choris-  
 CC mate pyruvate lyase (CPL) in pET24a Escherichia coli expression construct is  
 CC described in the examples of the present invention

XX Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;

Query Match 72.4%; Score 495; DB 6; Length 495;  
 Best Local Similarity 100.0%; Pred. No. 7.2e-156;  
 Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 ATGTACACCCCGCTTAAGCAACTGCGTGCCTCGCTATTGTAAAGAGATCCCTGCC 246  
 Db 1 ATGTACACCCCGCTTAAGCAACTGCGTGCCTCGCTATTGTAAAGAGATCCCTGCC 60  
 QY 247 CTGGATCCGCAACTGCTGCACCTGGCTGCTCGAGATTCCATGACAAACGTTTGA 306  
 Db 61 CTGGATCCGCAACTGCTGCACCTGGCTGCTCGAGATTCCATGACAAACGTTTGA 120  
 QY 307 CAGCAGGAGAAAACGGTAAGCGTGACGATGATCCGCGAAGGGTTTGTGAGCAGAATGA 366  
 Db 121 CAGCAGGAGAAAACGGTAAGCGTGACGATGATCCGCGAAGGGTTTGTGAGCAGAATGA 180  
 QY 367 ATCCCGAAGAACTGCGCTGCTGCCGAAGAGTCTGTTACTGTTAGTGAATTTTG 426  
 Db 181 ATCCCGAAGAACTGCGCTGCTGCCGAAGAGTCTGTTACTGTTAGTGAATTTTG 240  
 QY 427 TTATGTGCGGATGGTCAACCGTGGCTTGGCGTGCCTGCTGCTGTTCTGTGCAACGTTA 486  
 Db 241 TTATGTGCGGATGGTCAACCGTGGCTTGGCGTGCCTGCTGCTGTTCTGTGCAACGTTA 300  
 QY 487 AGCGGCGCGAGCTGCGCTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTC 546  
 Db 301 AGCGGCGCGAGCTGCGCTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTC 360  
 QY 547 ACATCATCGACATTAAACCGGAGCTTTATGAGATAGGCGGTGATCCGGGCTGTGGGG 606  
 Db 361 ACATCATCGACATTAAACCGGAGCTTTATGAGATAGGCGGTGATCCGGGCTGTGGGG 420  
 QY 607 CGACGTTCCCGCTGGATTAGCGGTAAACCGCTGCTGCTTAACAGAACTGTTTACCG 666  
 Db 421 CGACGTTCCCGCTGGATTAGCGGTAAACCGCTGCTGCTTAACAGAACTGTTTACCG 480  
 QY 667 GCGTCACCGTTGTAC 681  
 Db 481 GCGTCACCGTTGTAC 495

RESULT 14  
 ADC25984

ID ADC25984 standard; DNA; 495 BP.

XX  
 AC ADC25984;

XX 18-DEC-2003 (first entry)

XX Escherichia coli choris-  
 DE mate pyruvate lyase DNA.

XX UDP-glucosyltransferase; pHBA; p-hydroxybenzoic acid ester glucoside;  
 KW liquid crystal polymer; LCP; methylparaben; preservative; food;  
 KW cosmetic industry; ds; gene; ubiC; CPL; choris-  
 mate pyruvate lyase.

XX Escherichia coli.

XX Key Location/Qualifiers  
 CDS 1..495

FT /\*tag= a

FT /product= "Escherichia coli choris-  
 FT mate pyruvate lyase

FT /note= "No stop codon"

XX WO2003066836-A2.

XX 14-AUG-2003.

XX 06-FEB-2003; 2003WO-US005863.

XX 07-FEB-2002; 2002US-0355511P.

XX (DUPO ) DU PONT DE NEMOURS & CO E I.

XX Meyer K, Van Dyk DE, Viitanen PV;

XX WPI; 2003-767259/72.

XX P-PSDB; ADC25985.

XX New nucleic acid encoding UDP-glucosyltransferase, useful for preparing  
 PT cells that produce p-hydroxybenzoic acid glucose ester, also the new  
 PT enzymes.

XX Example 9; SEQ ID NO 37; 161pp; English.

XX The invention relates to a novel isolated nucleic acid that encodes a UDP  
 CC -glucosyltransferase. The method of the invention may be used to  
 CC transform microorganisms or green plant cells so that these produce a  
 CC higher level of high-volume chemicals or materials, such as pHBA (p-  
 CC hydroxybenzoic acid) ester glucoside which is an intermediate for pHBA, a  
 CC monomer for liquid crystal polymers (LCPs) and starting material for  
 CC methylparaben. Methylparaben is a preservative commonly used in the food  
 CC and cosmetic industries. The encoded enzymes may be used for in vitro  
 CC production of these compounds and for identifying similar enzymes by  
 CC sequence comparison. The current sequence is that of the Escherichia coli  
 CC choris-  
 CC mate pyruvate lyase (CPL; UbiC) DNA of the invention.

XX Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;

Query Match 72.4%; Score 495; DB 10; Length 495;  
 Best Local Similarity 100.0%; Pred. No. 7.2e-156;  
 Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 ATGTACACCCCGCTTAAGCAACTGCGTGCCTGCTATTGTAAAGAGATCCCTGCC 246  
 Db 1 ATGTACACCCCGCTTAAGCAACTGCGTGCCTGCTATTGTAAAGAGATCCCTGCC 60  
 QY 247 CTGGATCCGCAACTGCTGCACCTGGCTGCTCGAGATTCCATGACAAACGTTTGA 306  
 Db 61 CTGGATCCGCAACTGCTGCACCTGGCTGCTCGAGATTCCATGACAAACGTTTGA 120  
 QY 307 CAGCAGGAGAAAACGGTAAGCGTGACGATGATCCGCGAAGGGTTTGTGAGCAGAATGA 366  
 Db 121 CAGCAGGAGAAAACGGTAAGCGTGACGATGATCCGCGAAGGGTTTGTGAGCAGAATGA 180  
 QY 367 ATCCCGAAGAACTGCGCTGCTGCCGAAGAGTCTGTTACTGTTAGTGAATTTTG 426  
 Db 181 ATCCCGAAGAACTGCGCTGCTGCCGAAGAGTCTGTTACTGTTAGTGAATTTTG 240  
 QY 427 TTATGTGCGGATGGTCAACCGTGGCTTGGCGTGCCTGCTGCTGTTCTGTGCAACGTTA 486  
 Db 241 TTATGTGCGGATGGTCAACCGTGGCTTGGCGTGCCTGCTGCTGTTCTGTGCAACGTTA 300  
 QY 487 AGCGGCGCGAGCTGCGCTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTC 546  
 Db 301 AGCGGCGCGAGCTGCGCTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTC 360



[illegible]

RESULT 15	
AAS90335	AAS90335 standard; cDNA; 1207 BP.
XX	
XX	
XX	AAS90335;
XX	
XX	
XX	13-FEB-2002 (first entry)
XX	
XX	DNA encoding novel human diagnostic protein #26139.
DE	
XX	
XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	
XX	Homo sapiens.
OS	
XX	
XX	WO200175067-A2.
PN	
XX	
XX	11-OCT-2001.
XX	
XX	30-MAR-2001; 2001WO-US008631.
XX	
XX	
PR	31-MAR-2000; 2000US-00540217.
PR	23-AUG-2000; 2000US-00649167.
XX	
XX	(HYSE-) HYSEQ INC.
XX	
XX	
PI	Drmanac RT, Liu C, Tang YT;
XX	
XX	WPI; 2001-639362/73.
DR	P-PSDB; ABG26148.
DR	

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

PS Claim 1; SEQ ID NO 26139; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AA364197-AA394584 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [http://wipo.int/pub/published/pct\\_sequences](http://wipo.int/pub/published/pct_sequences)

XX	Sequence	1207 BP;	316 A;	265 C;	288 G;	338 T;	0 U;	0 Other;
	Query Match	69.1%	Score	472.8;	DB 5;	Length	1207;	
	Best Local Similarity	99.2%;	Pred. No.	3.6e-148;				
	Matches	496;	Conservative	0;	Mismatches	2;	Indels	2;
Qy	187	ATGTCACACCCGGTTAAGCGAACTGCGTGCCTCGCTATTTGTAAGAGATCCCTGCC	246					
Db	652	ATGTCACACCCGGTTAAGCGAACTGCGTGCCTCGCTATTTGTAAGAGATCCCTGCC	711					
Qy	247	CTGGATCCGCAACTGCTCGACTGGCTGTTCTCGAGGATTCATAGCAAAACGTTTGTAA	306					
Db	712	CTGGATCCGCAACTGCTCGACTGGCTGTTCTCGAGGATTCATAGCAAAACGTTTGTAA	771					
Qy	307	CAGCAGGAAAAACGGTAAGCGTGACGATATCGCGAAGGGTTGTCGAGCAGATGAA	366					
Db	772	CAGCAGGAAAAACGGTAAGCGTGACGATATCGCGAAGGGTTGTCGAGCAGATGAA	831					
Qy	367	ATCCCCGAAGAACTGCCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTTG	426					
Db	832	ATCCCCGAAGAACTGCCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTTG	891					
Qy	427	T-TATGTGCGATGGTGAAACGGTGGCTTGGC--GGTGTACCGTGGTTCCTGTCTCAACGT	484					
Db	892	TGATGTGCGATGGCGAAACCGCGGCTTGC CGGGT CGTACCGTGTCTGTGTCAACGT	951					
Qy	485	TAAGCGGCGGAGCTGGCCGTTACAAAAATTGGGTAAAAACCGCGTTAGGACGCTATCTGT	544					
Db	952	TAAGCGGCGGAGCTGGCGTTACAAAAATTGGGTAAAAACCGCGTTAGGACGCTATCTGT	1011					
Qy	545	TCATCATATCGAATTAAACCGGACCTTTATTGAGATAGGCCGTGATGCCGGGCTGTGGG	604					
Db	1012	TCATCATATCGAATTAAACCGGACCTTTATTGAGATAGGCCGTGATGCCGGGCTGTGGG	1071					
Qy	605	GGGACGCTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTTAAACAGAACTGTTTTTAC	664					
Db	1072	GGGACGCTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTTAAACAGAACTGTTTTTAC	1131					
Qy	665	CGGCGTCACCGTTGTACTAA	684					
Db	1132	CGGCGTCACCGTTGTACTAA	1151					

Search completed: May 29, 2006, 12:14:11  
Job time : 937 secs

GenCore version 5.1.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 29, 2006, 20:59:50 ; Search time 746.985 Seconds

(without alignments)  
2380.132 Million cell updates/sec

Title: US-10-718-311-16

Perfect score: 887

Sequence: 1 MQVWMSHPALQLRALRYC.....RLSGKPLLLTELFLPASPLY 170

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 5244920 seqs, 3486124231 residues

Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlp  
-Q-/abss/ABSSWEB spool/US10718311/runat\_26052006\_164914\_13428/app.query.fasta\_1  
-DB=N Geneseq -QFM=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human4.0.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss02p  
-USER=US10718311 @CN 1.1 2295 @runat\_26052006\_164914\_13428 -NCFU=6 -ICPU=3  
-NO WMAP -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPEXT=10 -XGAPEXT=0.5 -FGAPOPEXT=6 -FGAPOPEXT=7  
-YGAPOPEXT=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 8.1\*

1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002as.\*  
7: Geneseqn2002bs.\*  
8: Geneseqn2003as.\*  
9: Geneseqn2003bs.\*  
10: Geneseqn2003cs.\*  
11: Geneseqn2003ds.\*  
12: Geneseqn2004as.\*  
13: Geneseqn2004bs.\*  
14: Geneseqn2005s.\*  
15: Geneseqn2006s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	887	100.0	684	6	ABA91838 Chloropia
2	887	100.0	684	6	Abk47712 DNA seque
3	887	100.0	684	10	Adc25988 Tomato/Es

4	887	100.0	684	14	ADV91622	Tomato ru
5	882	97.2	3452	14	ADV91636	Adv91636 C. paraps
6	854	96.3	495	2	AAQ92409	E.coli ch
7	854	96.3	495	6	ABA91837	Escherich
8	854	96.3	495	6	ABK47709	DNA seque
9	854	96.3	495	10	ADC25984	Escherich
10	854	96.3	498	14	ADV91618	Escherich
11	854	96.3	854	4	AAD06956	ubc gene
12	854	96.3	6641	4	AAD06957	pHE2 comp
13	851	95.9	1971	14	ADV91643	C. paraps
14	851	95.9	2000	2	AAI29820	E. coli u
15	811	91.4	1207	5	AAS90335	DNA encod
16	791	89.2	1918	2	AAQ67671	E.coli ub
17	673	75.9	645	11	ACH95439	Klebselli
18	417	47.0	519	10	ACF70957	Photorhab
19	417	47.0	110000	10	ACF67367	40
20	417	47.0	110000	10	ACF65388	07
21	401	45.2	537	10	ADF02837	Bacterial
22	186	21.0	630	11	ABD08838	Pseudomon
23	174.5	19.7	519	12	ADL04203	DNA encod
24	174	19.6	100	8	ACD78861	E. coli K
25	171.5	19.3	269223	4	AAF28554	Genomic f
26	167	18.8	100	8	ACD78860	E. coli K
27	165	18.6	100	8	ACD78859	E. coli K
28	159	17.9	540	9	ADA29056	DNA encod
29	146.5	16.5	1248	11	ABD08757	Pseudomon
30	133.5	15.1	55839	12	ADN36893	X. albili
31	130.5	14.7	609	12	ADN36906	X. albili
32	121	13.6	195	10	ADF02801	Bacterial
33	97.5	11.0	475	6	AAS97252	Neisseria
34	90	10.1	2127	4	ABA95459	Thermus t
35	90	10.1	9133	3	AAA81730	N. mening
36	90	10.1	110000	3	AAA81490	07
37	90	10.1	349880	3	AAF21508	Neisseria
38	88.5	10.0	2892	15	AEF63232	Human UNC
39	87	9.9	42811	12	ADQ97948	Human can
40	87	9.8	59589	11	ACN44724	Mouse gen
41	86.5	9.8	110000	10	ACF67367	43
42	86.5	9.8	110000	10	ACF65388	04
43	86	9.7	522	10	ABZ33133	Continuation (5 of
44	85.5	9.6	2157	1	AAH70438	N. gonorr
45	85.5	9.6	2907	11	ADM02382	Streptoco

ALIGNMENTS

RESULT 1  
ABA91838  
ID ABA91838 standard; DNA; 684 BP.

AC ABA91838;

XX 29-AUG-2003 (revised)  
DT 15-MAY-2002 (first entry)

XX Chloroplast transit peptide-chorismate pyruvate lyase fusion gene.

XX Chloroplast transit peptide; tomato; Rubisco; plant;

KW ribulose-1,5-bisphosphate carboxylase; chorismate pyruvate lyase; CPl;  
KW enzyme; p-hydroxybenzoic acid; transgenic plant; gene; ds.

XX Lycopersicon esculentum.

OS Escherichia coli.

OS Chimeric.

FH Key Location/Qualifiers  
CDS 1..495  
FT /\*tag= a  
FT /product= "chloroplast-targeted CPL fusion"

FT transit\_peptide 1..186

FT /\*tag= b

FT mat\_peptide 187..492

FT note= "tomato Rubisco chloroplast targeting sequence"

```

FT      /*tag= C
FN      /note= "E. coli chorismate pyruvate lyase"
XX
XX      WO200194607-A2.
XX      13-DEC-2001.
XX
XX      22-MAY-2001; 2001WO-US016661.
XX
XX      02-JUN-2000; 2000US-0209854P.
XX      (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX      Meyer K, Van Dyk DE, Viitanen PV;
XX
XX      WPI; 2002-226795/28.
XX      P-PSDB; AAM50959.
XX
XX      Producing para-hydroxy benzoic acid in green plant, comprises expression
XX      of unique expression cassette containing gene encoding chorismate
XX      pyruvate lyase operably linked to specific chloroplast targeting
XX      sequence.
XX
XX      Claim 17; Page 57; 60pp; English.
XX
XX      The present nucleotide sequence is that of an open reading frame encoding
XX      chloroplast-targeted chorismate pyruvate lyase (CPL) fusion protein TP-
XX      CPL (see AAM50959). PCR (see ABA91841-42) was used to generate a DNA
XX      fragment corresponding to the transit peptide of tomato Rubisco small
XX      subunit and the first 4 amino acid residues of mature Rubisco. The DNA
XX      fragment was ligated into pET24a-CPL, which carries the Escherichia coli
XX      CPL open reading frame (see ABA91837). The TP-CPL construct is an example
XX      of expression cassettes of the invention that are designed for the high-
XX      level production of p-hydroxybenzoic acid (pHBA) in higher plants.
XX      Cleavage of TP-CPL in the chloroplast releases a novel polypeptide (see
XX      AAM50961) that has full enzyme activity, converting chorismate to pHBA. A
XX      claimed plant comprising a CPL expression cassette is selected from
XX      soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat,
XX      barley, oats, sorghum, rice, Arabidopsis, sugarcane, canola,
XX      millet, bean, pea, rye, flax or a forage grass. pHBA is a monomeric
XX      component of liquid crystal polymers which have application in the
XX      automotive, electrical and other industries. (Updated on 29-AUG-2003 to
XX      standardise OS field)
XX
XX      SQ      Sequence 684 BP; 160 A; 171 C; 179 G; 174 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      2,04e-92      Length:      684
Score:          887.00      Matches:      170
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:      100.0%      Indels:      0
DB:              6      Gaps:        0

US-10-718-311-16 (1-170) x ABA91838 (1-684)

QY      1      MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20
DB      172      ATGCAGGTGTGCATATGTCACACCCCGCTTAACGCAACTGCGTGCCTGCCTATTGT 231

QY      21      LysGluIleProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40
DB      232      AAAGAGATCCCTGCCCTGGATCCGCAACTGCTCGACTGGCTGTCTGGAGATTCCATG 291

QY      41      ThrLysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPhe 60
DB      292      ACAAAACGTTTGAACACAGCAGGAAAACGGTAACGGTGACGATGATCCGCGAAGGGTT 351

QY      61      ValGluGlnAsnGluIleProGluGlnLeuProLeuLeuProLysGluSerArgTyrTrp 80
DB      352      GTCGACGAGAATGAATCCCGCAAGAACTCCGCGCTGCTGCCGAAAGAGTCTCGTTACTGG 411

QY      81      LeuArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100

```

---

```

Db      412      TTACGTGAATTTTGTATGTGCCGATGGTGAAACCGTGGCTTGCCTGTCGACCGTCGTT 471
QY      101      ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120
DB      472      CCTGTGTCAACGTTAAGCGGCGCGAGCTGGCGTTACAAAAATTGGGTAAACGCGGTTA 531
QY      121      GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAsp 140
DB      532      GGACGCTATCTCTGTTCAATCATCATGACATTAAACCCGGGACTTTATTGAGATAGGCGGTGAT 591
QY      141      AlaGlyLeuTrpGlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 160
DB      592      GCCGGGCTGTGGGGCGACGTTCCGCTCGCATTAACGGGTAAACCGCTGTGTGTAACA 651
QY      161      GluLeuPheLeuProAlaSerProLeuTyr 170
DB      652      GNACTGTTTTTACCGGCGTCACCGTTGTAC 681

RESULT 2
ABK47712
ID      ABK47712 standard; DNA; 684 BP.
XX
XX      AC      ABK47712;
XX
XX      DT      18-JUN-2002 (first entry)
XX      DE      DNA sequence of ORF for TP-CPL fusion protein.
XX      KW      SNG1; sinapoylglucose accumulator 1; SMT; aromatic ester; plant;
XX      KW      sinapoylglucose;malate sinapoyltrifluoromethyltransferase; glycosylated aromatic acid;
XX      KW      malate conjugated aromatic acid; polymer synthesis; ubiC; TP-CPL; gene;
XX      KW      carboxylic acid conjugated aromatic acid; chorismate pyruvate lyase;
XX      KW      tomato Rubisco small subunit precursor; rbcS2; mutant; ds.
XX      OS      Escherichia coli.
XX      OS      Lycopersicon esculentum.
XX      OS      Synthetic.
XX      OS      Chimeric.
XX      FH      Key      Location/Qualifiers
XX      FT      CDS      1..684
XX      FT      /*tag= a
XX      FT      /product= "TP-CPL fusion protein"
XX
XX      PN      WO200204653-A2.
XX
XX      PD      17-JAN-2002.
XX
XX      PF      05-JUL-2001; 2001WO-US021283.
XX
XX      PR      07-JUL-2000; 2000US-0216615P.
XX
XX      PA      (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX      PI      Flint D, Meyer K, Viitanen PV;
XX
XX      WI      WPI; 2002-303779/34.
XX      DR      P-PSDB; AAU77942.
XX
XX      PT      Producing aromatic acid conjugates, involves contacting glycosylated
XX      aromatic acid with malate, alpha-hydroxycarboxylic acid or alcohol in the
XX      presence of sinapoylglucose:malate sinapoyltrifluoromethyltransferase.
XX
XX      Example; Page 70; 72pp; English.
XX
XX      The present invention relates to the isolation of Arabidopsis thaliana
XX      gene (SNG1, sinapoylglucose accumulator 1) encoding
XX      sinapoylglucose:malate sinapoyltrifluoromethyltransferase (SMT). SMT catalyses the
XX      substitution of a glucose moiety on a glycosylated aromatic acid with a
XX      malate moiety to form a malate conjugated aromatic acid. The enzyme is
XX      useful for producing malate conjugated aromatic acids, carboxylic acid
XX      conjugated aromatic acids or aromatic esters. The malate conjugated

```

CC aromatic acids are useful in the synthesis of various polymers. The  
 CC present sequence representing the ORF (open reading frame) for TP-CPL  
 CC (tomato Rubisco small subunit precursor for rbcS2-Escherichia coli  
 CC chorismate pyruvate lyase (CPL)) fusion protein is described in the  
 CC examples of the present invention

XX Sequence 684 BP; 160 A; 171 C; 179 G; 174 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 2,04e-92 Length: 684  
 Score: 887.00 Matches: 170  
 Percent Similarity: 100.0% Conservatives: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 100.0% Indels: 0  
 DB: 6 Gaps: 0

US-10-718-311-16 (1-170) x ABK47712 (1-684)

QY 1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20  
 DB 172 ATGCAGGTGGCATATGTACACCCCGCGTTAAACGCAACTGCGTGGCTATTGT 231  
 QY 21 LysGluLeuProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40  
 DB 232 AAAGAGATCCCTGCCCTGGATCCGCAACTGCTCGACTGGCTGTGTGGAGGATTCATG 291  
 QY 41 ThrLysArgPheGluGlnGlnLysThrValSerValThrMetLeuArgGluGlyPhe 60  
 DB 292 ACANAAACGTTTTGAACGACGAGGNAACGGTAAGCGTGACGATATCCGGAAGGGTTT 351  
 QY 61 ValGluGlnAsnGluLeuProGluGluLeuProLeuLeuProLysGluSerArgTyrTrp 80  
 DB 352 GTCCAGCAGCAATGAATATCCCGAAGAACTGCCGCTCTCCGAAAGAGTCTCGTTACTG 411  
 QY 81 LeuArgGluLeuLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100  
 DB 412 TTACGTGAAATTTTGTATGTCCGATGGTGGAACCGTGGCTTGCCTGCGTACCGTCGT 471  
 QY 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120  
 DB 472 CTGTGTCAACGTTAAGCGGGCGGAGCTGGCGTTACAAAAATTTGGGTAAACCGCGTTA 531  
 QY 121 GlyArgTyrLeuPheThrSerThrLeuThrArgAspPheileGluLeuGlyArgAsp 140  
 DB 532 GCAGCGTATCTGTTCACATCATCGACATTAAACCGGGAGCTTTTATTGAGATAGCGCGTAT 591  
 QY 141 AlaGlyLeuTrpGlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 160  
 DB 592 GCCGGGCTGTGGGGCGAGCTTCCCGCTCGCATTAAGCGGTAAACCGCTGTGTCTAACA 651  
 QY 161 GluLeuPheLeuProAlaSerProLeuTyr 170  
 DB 652 GAACGTGTTTTTACCGGCGTACCGCTTGATAC 681

#### RESULT 3

ADC25988

ID ADC25988 standard; DNA; 684 BP.

XX AC ADC25988;

XX AC ADC25988;

DT 18-DEC-2003 (first entry)

DE Tomato/Escherichia coli chimeric TP-CPL DNA.

KW UDP-glucosyltransferase; pABA; p-hydroxybenzoic acid ester glucoside;  
 KW liquid crystal polymer; LCP; methylparaben; preservative; food;  
 KW cosmetic industry; ds; gene; tomato; TP-CPL; chimeric;  
 KW chorismate pyruvate lyase; plant.

XX Chimeric.

OS Lycopersicon esculentum.

OS Escherichia coli.

XX

#### Location/Qualifiers

1..684  
 /tags= a  
 /product= "Tomato/Escherichia coli chimeric TP-CPL  
 protein"

XX WO2003066836-A2.

XX 14-AUG-2003.

XX 06-FEB-2003; 2003WO-US0005863.

XX 07-FEB-2002; 2002US-0355511P.

XX (DUPO ) DU PONT DE NEMOURS & CO E I.

XX Meyer K, Van Dyk DE, Viitanen PV;

XX WPI; 2003-767259/72.

XX P-PSDB; ADC25989.

XX New nucleic acid encoding UDP-glucosyltransferase, useful for preparing  
 cells that produce p-hydroxybenzoic acid glucose ester, also the new  
 enzymes.  
 XX Example 9; SEQ ID NO 41; 161pp; English.  
 XX The invention relates to a novel isolated nucleic acid that encodes a UDP  
 -glucosyltransferase. The method of the invention may be used to  
 transform microorganisms or green plant cells so that these produce a  
 higher level of high-volume chemicals or materials, such as pABA (p-  
 hydroxybenzoic acid) ester glucoside which is an intermediate for pABA, a  
 monomer for liquid crystal polymers (LCPs) and starting material for  
 methylparaben. Methylparaben is a preservative commonly used in the food  
 and cosmetic industries. The encoded enzymes may be used for in vitro  
 production of these compounds and for identifying similar enzymes by  
 sequence comparison. The current sequence is that of the  
 tomato/Escherichia coli chimeric TP-CPL (chorismate pyruvate lyase) DNA  
 of the invention.

XX Sequence 684 BP; 160 A; 171 C; 179 G; 174 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 2,04e-92 Length: 684  
 Score: 887.00 Matches: 170  
 Percent Similarity: 100.0% Conservatives: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 100.0% Indels: 0  
 DB: 10 Gaps: 0

US-10-718-311-16 (1-170) x ADC25988 (1-684)

QY 1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20  
 DB 172 ATGCAGGTGGCATATGTACACCCCGCGTTAAACGCAACTGCGTGGCTATTGT 231  
 QY 21 LysGluLeuProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40  
 DB 232 AAAGAGATCCCTGCCCTGGATCCGCAACTGCTCGACTGGCTGTGTGGAGGATTCATG 291  
 QY 41 ThrLysArgPheGluGlnGlnLysThrValSerValThrMetLeuArgGluGlyPhe 60  
 DB 292 ACANAAACGTTTTGAACGACGAGGNAACGGTAAGCGTGACGATATCCGGAAGGGTTT 351  
 QY 61 ValGluGlnAsnGluLeuProGluGluLeuProLeuLeuProLysGluSerArgTyrTrp 80  
 DB 352 GTCCAGCAGCAATGAATATCCCGAAGAACTGCCGCTCTCCGAAAGAGTCTCGTTACTG 411  
 QY 81 LeuArgGluLeuLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100  
 DB 412 TTACGTGAAATTTTGTATGTCCGATGGTGGAACCGTGGCTTGCCTGCGTACCGTCGT 471  
 QY 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120

```

Db 472 CCTGTCTCAACGTTAAGCGCGCGGAGCTGGCTTACAAAAATTGGGTAAACGCCGTTA 531
Qy 121 GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAsp 140
Db 532 GGACGCTATCTGTTACATCATCATGACATTAACCGGGACTTTATTGAGTAGGCGGTGAT 591
Qy 141 AlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 160
Db 592 GCCGGCTGTGGGGCGGACGTTCCCGCTGGGATTAAGCGGTAAACCGCTGTTGTAACA 651
Qy 161 GluLeuPheLeuProAlaSerProLeuTyr 170
Db 652 GAACTGTTTTTACCGGCGTCACCGTTGTAC 681

```

## RESULT 4

ADV91622

ID ADV91622 standard; DNA; 684 BP.

XX AC

ADV91622;

XX DT 10-MAR-2005 (first entry)

XX DE Tomato rubisco small subunit precursor TP-E. coli CPL chimeric DNA.

XX KW Hydroquinone glucoside; arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside;  
 KW chorismate pyruvate lyase; CPL; 4-hydroxycinnamoyl-CoA hydratase/lyase;  
 KW HCHL; gene expression; hydroquinone; antiseptic; transgenic plant;  
 KW genetically engineered microorganism; antioxidant; antimicrobial agent;  
 KW anti-inflammatory agent; carcinogenic inhibitor; melanoma; cytostatic;  
 KW neoplasm; pHBA; para-hydroxybenzoic acid; p-hydroxybenzoic acid;  
 KW rubisco small subunit precursor; transit peptide;  
 KW chloroplast transit peptide; TP; chloroplast; gene; ds.

XX OS Lycopersicon esculentum.

OS Escherichia coli.

XX OS Chimeric.

XX FH Key Location/Qualifiers  
 XX CDS 1..684

FT /\*tag= a

FT /product= "Tomato rubisco small subunit precursor TP-E.  
 FT coll CPL fusion protein"

FT FT

FT FT

XX PN US2004261147-A1.

XX PD 23-DEC-2004.

XX PF 16-JUN-2003; 2003US-00462162.

XX PR 16-JUN-2003; 2003US-00462162.

XX PA (MEYE/) MEYER K.

PA (VIIT/) VIITANEN P V.

PA (FLIN/) FLINT D.

XX XX

XX PI Meyer K, Viitanen PV, Flint D;

XX XX

XX DR WPI; 2005-057232/06.

XX DR P-PSDB; ADV91623.

XX XX

XX PT Producing hydroquinone glucoside in a green plant comprises growing a  
 XX green plant having nucleic acid fragments and chorismate pyruvate lyase  
 XX expression cassette.

XX PS Example 4; SEQ ID NO 33; 70pp; English.

XX XX

XX CC The present invention relates to methods and materials to produce  
 CC hydroquinone glucoside (arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside)  
 CC in genetically modified green plants and microorganisms. The method  
 CC relies upon transgenic plants or genetically modified microorganisms that  
 CC produce increased levels of the initial substrate para-hydroxybenzoic  
 CC acid (p-hydroxybenzoic acid; pHBA) in a biosynthetic pathway to produce

CC arbutin. Plants and microbes can be genetically engineered to produce  
 CC high levels of pHBA either by functional expression of the bacterial  
 CC protein chorismate pyruvate lyase (CPL) or by the expression of 4-  
 CC hydroxycinnamoyl-CoA hydratase/lyase (HCHL) or a combination of both.  
 CC Arbutin is useful as an antioxidant, antimicrobial agent, anti-  
 CC inflammatory agent and inhibitor of carcinogenesis (melanoma). The  
 CC present sequence is tomato rubisco small subunit precursor (rbcS2)  
 CC transit peptide (chloroplast transit peptide; TP)-Escherichia coli CPL  
 CC chimeric DNA. This sequence is present in the E. coli expression vector  
 XX construct pET24a-TP-CPL.

SQ Sequence 684 BP; 160 A; 171 C; 179 G; 174 T; 0 U; 0 Other;

## Alignment Scores:

Pred. NO.: 2.04e-32 Length: 684  
 Score: 887.00 Matches: 170  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 100.0% Indels: 0  
 DB: 14 Gaps: 0

US-10-718-311-16 (1-170) x ADV91622 (1-684)

Qy 1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20  
 Db 172 ATGCAGGTGTGGCATATGTACACCCCGCGTTAACGCAACTGCGTGCCTGCGCTATTGT 231

Qy 21 LysGluIleProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMet 40  
 Db 232 AAAGAGATCCCTGCGCTGGATCCGCAACTGCTCAGCTGGCTGTTGCTGGAGGATTCATG 291

Qy 41 ThrLysArgPheGluGlnGlnGlyThrValSerValThrMetIleArgGluGlyPhe 60  
 Db 292 ACAAAACGTTTTGAACAGCAGCGGAGGAAAAACGGTAAGCGTGACGATGCCGGAAGGGTTT 351

Qy 61 ValGluGlnGlnGluIleProGluGluLeuProGluLeuProLysGluSerArgTyrTrp 80  
 Db 352 GTCGAGCAGAGATGAATCCCGAAGAACTGCCGCTGCTGCCGAAAGAGTCTCGTTACTGG 411

Qy 81 LeuArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100  
 Db 412 TTACGTGAATTTTGTATTATGTGCGGATGGTGAACCGTGGCTTGCCTGCGTACGTCGT 471

Qy 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120  
 Db 472 CCTGTGCAACGTTAAGCGGCGGAGCTGCGGTACAAAAATTGGGTAAACGCCGTTA 531

Qy 121 GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAsp 140  
 Db 532 GGACGCTATCTGTTACATCATCATGACATTAACCGGGAGCTTTATTGAGTAGGCGGTGAT 591

Qy 141 AlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 160  
 Db 592 GCCGGCTGTGGGGCGGACGTTCCCGCTGGGATTAAGCGGTAAACCGCTGTTGTAACA 651

Qy 161 GluLeuPheLeuProAlaSerProLeuTyr 170  
 Db 652 GAACTGTTTTTACCGGCGTCACCGTTGTAC 681

## RESULT 5

ADV91636

ID ADV91636 standard; DNA; 3452 BP.

XX AC

ADV91636;

XX XX

XX DT 10-MAR-2005 (first entry)

XX XX

XX DE C. parapsilosis PHB 1-H- E. coli CPL- A. thaliana UGT72B1 chimeric DNA.  
 XX KW Hydroquinone glucoside; arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside;  
 KW chorismate pyruvate lyase; CPL; 4-hydroxycinnamoyl-CoA hydratase/lyase;  
 KW HCHL; gene expression; hydroquinone; antiseptic; transgenic plant;  
 KW genetically engineered microorganism; antioxidant; antimicrobial agent;  
 KW

KW anti-inflammatory agent; carcinogenic inhibitor; melanoma; cytostatic;  
KW neoplasm; p-HBA 1-hydroxylase; p-HBA; para-hydroxybenzoic acid;  
KW p-hydroxybenzoic acid; UDP-glucosyltransferase; UGT72B1; gene; ds.  
XX Arabidopsis thaliana.  
OS Candida parapsilosis.  
OS Escherichia coli.  
OS Chimeric.

Key Location/Qualifiers  
CDS 4..501  
FT /tag= a  
FT /product= "Escherichia coli CPL protein"  
FT 541..1980  
FT /\*tag= b  
FT /product= "Candida parapsilosis p-HBA 1-H protein"  
FT /transl\_except= (pos: 559..561, aa:Trp)  
CDS 2004..3446  
FT /\*tag= c  
FT /product= "Arabidopsis thaliana UGT72B1 protein"

US2004261147-A1.

23-DEC-2004.

16-JUN-2003; 2003US-00462162.

16-JUN-2003; 2003US-00462162.

(MEYE/) MEYER K.

(VIIT/) VIITANEN P V.

(FLIN/) FLINT D.

Meyer K, Viitanen PV, Flint D;

WPI; 2005-057232/06.

P-PSDB; ADV91605, ADV91619, ADV91631.

Producing hydroquinone glucoside in a green plant comprises growing a green plant having nucleic acid fragments and chorismate pyruvate lyase expression cassette.

Example 7; SEQ ID NO 47; 70pp; English.

The present invention relates to methods and materials to produce hydroquinone glucoside (arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside) in genetically modified green plants and microorganisms. The method relies upon transgenic plants or genetically modified microorganisms that produce increased levels of the initial substrate para-hydroxybenzoic acid (p-hydroxybenzoic acid; p-HBA) in a biosynthetic pathway to produce arbutin. Plants and microbes can be genetically engineered to produce high levels of p-HBA either by functional expression of the bacterial protein chorismate pyruvate lyase (CPL) or by the expression of 4-hydroxycinnamoyl-CoA hydratase/lyase (HCHL) or a combination of both. Arbutin is useful as an antioxidant, antimicrobial agent, anti-inflammatory agent and inhibitor of carcinogenesis (melanoma). The present sequence is Candida parapsilosis p-HBA 1-hydroxylase (p-H 1-H) - Escherichia coli chorismate pyruvate lyase (CPL) and Arabidopsis thaliana UDP-glucosyltransferase (UGT72B1) chimeric DNA.

Sequence 3452 BP; 932 A; 730 C; 856 G; 934 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1,48e-88	Length:	3452
Score:	862.00	Matches:	166
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	97.2%	Indels:	0
DB:	14	Gaps:	0

US-10-718-311-16 (1-170) x ADV91636 (1-3452)

Qy 5 HisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIlePro 24

Db	1	CATATGTCACCCCGCTTAACGCAACTGGGTGCGCTATTGTTAAAGAGATCCCT	60
Qy	25	AlaLeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPhe	44
Db	61	GGCTGGATCCGCACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTT	120
Qy	45	GluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsn	64
Db	121	GAACAGCAGGAAACGCTAGCGTACGATGATCCGCGAGGGTTTGTCTGAGCAGAAAT	180
Qy	65	GluIleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIle	84
Db	181	GAATATCCCGAAGAACTGCTCGCTGCTCCGAAAGAGTCTCGTTACTGGTTACGTAAAT	240
Qy	85	LeuLeuCysAlaAspGlyClnProTrpLeuAlaGlyArgThrValValProValSerThr	104
Db	241	TTGTTATGTGCGGATGGTGAACCGTGGCTTGGCGGTGCTACCGTCTCTGTGTCAACG	300
Qy	105	LeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeu	124
Db	301	TTAAGCGGCGGAGCTGGCTTACAAAATTTGGGTAAACCGCTTAGGACGCTATCTG	360
Qy	125	PheThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrp	144
Db	361	TTACATCATCGACATTAAACCGGACTTTATTAGATAGGCGGTGATGCGGGCTGTGG	420
Qy	145	GlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeu	164
Db	421	GGCGACGTTCCCGCTCGATTAAACCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTA	480
Qy	165	ProAlaSerProLeuTyr 170	
Db	481	CCGCGCTACCGCTTGTAC 498	

RESULT 6

AAQ92409

ID AAQ92409 standard; DNA; 495 BP.

AC AAQ92409;

XX

DT 11-JAN-1996 (first entry)

XX

DE E.coli chorismate-pyruvate lyase coding sequence.

XX

KW Chorismate-pyruvate lyase gene; ubiC gene; transgenic plant;

KW tobacco mosaic virus resistance; antiviral activity;

KW p-hydroxybenzoic acid; ss.

XX Escherichia coli.

OS

PN DE4423022-C1.

XX

PD 24-MAY-1995.

XX

PF 30-JUN-1994; 94DE-04423022.

XX

PR 30-JUN-1994; 94DE-04423022.

XX

PA (HEID/) HEIDE L.

XX

PI Heide L, Siebert M, Severin K;

XX

DR WPI: 1995-186908/25.

DR P-PSDB; AAR74742.

XX

PT Transgenic plants with increased content of resistance factor - contg.

PT bacterial DNA coding for enzyme catalysing prodn. of resistance factor,

PT e.g. antiviral p-hydroxy-benzoic acid.

XX

PS Claim 3; Page 5; 7pp; German.

CC Transgenic plants containing a bacterial gene which encodes an enzyme

```
CC able to catalyze prodn. of an antiviral, bactericidal, fungicidal or
CC insecticidal factor are new. A preferred gene is ubiC from E.coli
CC (AAQ92409) which codes for chorismate-pyruvate lyase (AAR74742). The
CC lyase catalyzes conversion of chorismate to p-hydroxybenzoic acid and
CC transgenic tobacco plants which express the active enzyme are resistant
CC to tobacco mosaic virus
XX
SQ Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      8.33e-89      Length:      495
Score:          854.00      Matches:    165
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    96.3%      Indels:     0
DB:             2      Gaps:       0

US-10-718-311-16 (1-170) x AAQ92409 (1-495)

Qy      6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 25
Db      1 ATGTCACACCCCGCGTTAACGCACTGCGTGCCTATTTGTAAGAAGATCCCTGCC 60
Qy      26 LeuAspProGlnLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db      61 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGA 120
Qy      46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
Db      121 CAGCAGGGAAGAAACGGTAAGCGTGACGATGATCCGGAAGGTTTGTTCGACGAGATGAA 180
Qy      66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeu 85
Db      181 ATCCCGAAGAACTGCGCTGTCGCGAAGAGTCTCGTTACTGTTACGTGAAATTTG 240
Qy      86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 105
Db      241 TTATGTCCGATGGTAACCGTGGTGTGCCGGTGTACCGTCTGTCTGTCAACGTTA 300
Qy      106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
Db      301 AGCGGCGGAGCTGGCGTTACAAAATTGGTAAACGCCGTTAGGACGCTATCTGTT 360
Qy      126 ThrSerSerThrLeuThrArgAspPheIleGluLeuGlyArgAspAlaGlyLeuTrpGly 145
Db      361 ACATCATCGACATTAAACCGGACTTTATTGAGATAGCCGCGTATGCGCGCTGTGGGG 420
Qy      146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
Db      421 CGACGTTCCCGCTGCGATTAGCGGTAAACCGCTGTTGCTAACGAACCTGTTTACCG 480
Qy      166 AlaSerProLeuTyr 170
Db      481 GCGTCACCGTTGTAC 495

RESULT 7
ABA91837
ID ABA91837 standard; DNA; 495 BP.
XX
AC ABA91837;
XX
DT 15-MAY-2002 (first entry)
XX
DE Escherichia coli chorismate pyruvate lyase gene coding region.
XX
KW Chorismate pyruvate lyase; CPL; enzyme; p-hydroxybenzoic acid;
XX transgenic plant; ubiC; gene; ds.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 1..495
FT /*tag= a
```

```
FT /partial
FT /product= "Chorismate_pyruvate_lyase"
XX /note= "the CDS does not include a stop codon"
PN WO200194607-A2.
XX
PD 13-DEC-2001.
XX
PF 22-MAY-2001; 2001WO-US016661.
XX
PR 02-JUN-2000; 2000US-0209854P.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI Meyer K, Van Dyk DE, Viitanen PV;
XX
WI; 2002-226795/28.
XX
P-PSDB; AAM50958.
XX
XX Producing para-hydroxy benzoic acid in green plant, comprises expression
XX of unique expression cassette containing gene encoding chorismate
XX pyruvate lyase operably linked to specific chloroplast targeting
XX sequence.
XX
XX Claim 4; Page 55; 60pp; English.
XX
XX The present sequence is that of the coding region of the Escherichia coli
XX strain W3110 ubiC gene in expression construct pET24a. The gene encodes
XX chorismate pyruvate lyase (CPL, see AAM50968). It was obtained by PCR
XX amplification of strain W3110 genomic DNA using primers (see ABA91839-40)
XX based on the published E. coli ubiC gene. The CPL open reading frame is
XX used in an expression cassette designed for the high-level production of
XX p-hydroxybenzoic acid (pHBA) in green plants. The expression cassette
XX comprises the CPL coding sequence operably linked to a promoter capable
XX of driving protein expression in higher plants. The cassette also has a
XX sequence encoding a chloroplast transit peptide, its natural cleavage
XX site, and a small portion of a transit peptide donor protein fused to the
XX N-terminus of CPL. The chloroplast targeting sequence targets the foreign
XX protein to the chloroplast and aids in its uptake into the organelle. The
XX cleavage site is unique to the transit peptide, and cleavage of the
XX chimeric protein encoded by the cassette at this site releases a novel
XX polypeptide that has full enzyme activity, comprising the mature CPL
XX enzyme and a small portion of transit peptide donor (see AAM50961). A
XX plant comprising the CPL expression cassette is claimed, and may be
XX soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat,
XX barley, oats, sorghum, rice, Arabidopsis, sugarbeet, sugarcane, canola,
XX millet, bean, pea, rye, flax or a forage grass. pHBA is a monomeric
XX component of liquid crystal polymers which have application in the
XX automotive, electrical and other industries
XX
SQ Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      8.33e-89      Length:      495
Score:          854.00      Matches:    165
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    96.3%      Indels:     0
DB:             6      Gaps:       0

US-10-718-311-16 (1-170) x ABA91837 (1-495)

Qy      6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 25
Db      1 ATGTCACACCCCGCGTTAACGCAACTGCGTGCCTATTTGTAAGAAGATCCCTGCC 60
Qy      26 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db      61 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGA 120
Qy      46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
Db      121 CAGCAGGGAAGAAACGGTAAGCGTGACGATGATCCGGAAGGTTTGTTCGACGAGATGAA 180
```



QY 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85  
Db 181 ATCCCCGAAGAACTGCGCTGCTGCGAAGAGCTCTCGTTACTGGTTACGTGAATTTTG 240  
QY 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 105  
Db 241 TTATGTGCCGATGGTGAACCGTGGCTTGGCGGTCGTACCGTCTTCCGTGTCAACGTTA 300  
QY 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125  
Db 301 ACGGGCCGAGCTGCGCTTACAAAATTGGGTAAACCGCGTTAGACGCTATCTGTTTC 360  
QY 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145  
Db 361 ACATCATCGACATTAAACCGGACTTTATTGAGATAGGCGGTGATCGCGGCTGTGGGG 420  
QY 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165  
Db 421 CGACGTTCCCGCTCGCATTAAGCGGTAAACCGCTGTGCTAACAGAACTGTTTTTACCG 480  
QY 166 AlaSerProLeuTyr 170  
Db 481 GCGTCACCGTTGTAC 495

RESULT 8

ID ABK47709 standard; DNA; 495 BP.

XX AC ABK47709;

XX 18-JUN-2002 (first entry)

XX DNA sequence of ORF for CPL in pET24a E. coli expression construct.

XX SNG1; sinapoylglucose accumulator 1; SMT; aromatic ester;  
KW sinapoylglucose:malate sinapoyltransferase; glycosylated aromatic acid;  
KW malate conjugated aromatic acid; polymer synthesis; ubiC; CPL; gene;  
KW carboxylic acid conjugated aromatic acid; chorismate pyruvate lyase;  
KW pET24a; ds.  
XX Escherichia coli.

XX Key Location/Qualifiers

FT 1. .495

FT /\*tag= a

FT /partial

FT /product= "CPL"

FT /note= "This sequence lacks a stop codon"

XX WO200204653-A2.

XX 17-JAN-2002.

XX 05-JUL-2001; 2001WO-US021283.

XX 07-JUL-2000; 2000US-0216615P.

XX (DUPO ) DU PONT DE NEMOURS & CO E I.

XX Flint D, Meyer K, Viitanen PV;

XX WPI; 2002-303779/34.

XX P-PSDB; AAU77941.

XX Producing aromatic acid conjugates, involves contacting glycosylated  
PT aromatic acid with malate, alpha-hydroxycarboxylic acid or alcohol in the  
PT presence of sinapoylglucose:malate sinapoyltransferase.  
XX Example; Page 68; 72pp; English.  
XX The present invention relates to the isolation of Arabidopsis thaliana  
CC gene (SNG1, sinapoylglucose accumulator 1) encoding

CC sinapoylglucose:malate sinapoyltransferase (SMT). SMT catalyses the  
CC substitution of a glucose moiety on a glycosylated aromatic acid with a  
CC malate moiety to form a malate conjugated aromatic acid. The enzyme is  
CC useful for producing malate conjugated aromatic acids, carboxylic acid  
CC conjugated aromatic acids or aromatic esters. The malate conjugated  
CC aromatic acids are useful in the synthesis of various polymers. The  
CC present sequence representing the ORF (Open reading frame) for chorismate  
CC pyruvate lyase (CPL) in pET24a Escherichia coli expression construct is  
CC described in the examples of the present invention  
XX  
SQ Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8.33e-89 Length: 495  
Score: 854.00 Matches: 165  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 96.3% Indels: 0  
DB: 6 Gaps: 0

US-10-718-311-16 (1-170) x ABK47709 (1-495)

QY 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25  
Db 1 ATGTCACACCCCGCGTTAACGCAACTGCGTGGCGCTATTGTAAAGAGATCCCTGCC 60  
QY 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45  
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCCATGACAAACGTTTTGAA 120  
QY 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65  
Db 121 CAGCAGGGGAAAAACGTTAAGCGTGACGATGATCCGCGAAGGGTTTCTCGAGCAGAAATGAA 180  
QY 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85  
Db 181 ATCCCGCAGAAACTGCGCGTCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTTG 240  
QY 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 105  
Db 241 TTATGTGCCGATGGTGAACCGTGGCTTGGCGGTCGTACCGTCTTCTGTGTCAACGTTA 300  
QY 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125  
Db 301 ACGGGCCGAGCTGCGCTTACAAAATTGGGTAAACCGCGTTAGACGCTATCTGTTTC 360  
QY 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145  
Db 361 ACATCATCGACATTAAACCGGACTTTATTGAGATAGGCGGTGATCGCGGCTGTGGGG 420  
QY 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165  
Db 421 CGACGTTCCCGCTCGCATTAAGCGGTAAACCGCTGTGCTAACAGAACTGTTTTTACCG 480  
QY 166 AlaSerProLeuTyr 170  
Db 481 GCGTCACCGTTGTAC 495

RESULT 9

ID ADC25984

XX ADC25984 standard; DNA; 495 BP.

XX AC ADC25984;

XX 18-DEC-2003 (first entry)

XX Escherichia coli chorismate pyruvate lyase DNA.

XX UDP-glucosyltransferase; PHBA; p-hydroxybenzoic acid ester glucoside;  
KW liquid crystal polymer; LCP; methylparaben; preservative; food;  
KW cosmetic industry; ds; gene; ubiC; CPL; chorismate pyruvate lyase.  
XX Escherichia coli.

```

XX FH Key Location/Qualifiers
XX FT 1. .495
XX FT /*tag= a
XX FT /partial
XX FT /product= "Escherichia coli chorismate pyruvate lyase
XX FT protein"
XX FT /note= "No stop codon"
XX
XX PN WO2003066836-A2.
XX
XX PD 14-AUG-2003.
XX
XX PF 06-FEB-2003; 2003WO-US005863.
XX
XX PR 07-FEB-2002; 2002US-0355511P.
XX
XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX PI Meyer K, Van Dyk DE, Viitanen PV;
XX
XX DR WPI; 2003-767259/72.
XX
XX DR P-PSDB; ABC25985.
XX
XX PT New nucleic acid encoding UDP-glucosyltransferase, useful for preparing
XX FT cells that produce p-hydroxybenzoic acid glucose ester, also the new
XX FT enzymes.
XX
XX PS Example 9; SEQ ID NO 37; 161pp; English.
XX
XX CC The invention relates to a novel isolated nucleic acid that encodes a UDP
XX CC -glucosyltransferase. The method of the invention may be used to
XX CC transform microorganisms or green plant cells so that these produce a
XX CC higher level of high-volume chemicals or materials, such as pHBA (p-
XX CC hydroxybenzoic acid) ester glucoside which is an intermediate for pHBA, a
XX CC monomer for liquid crystal polymers (LCPs) and starting material for
XX CC methylparaben. Methylparaben is a preservative commonly used in the food
XX CC and cosmetic industries. The encoded enzymes may be used for in vitro
XX CC production of these compounds and for identifying similar enzymes by
XX CC sequence comparison. The current sequence is that of the Escherichia coli
XX CC chorismate pyruvate lyase (CPL; UbiC) DNA of the invention.
XX
XX SQ Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 8,33e-89 Length: 495
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.3% Indels: 0
DB: 10 Gaps: 0

US-10-718-311-16 (1-170) x ABC25984 (1-495)

QY 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
DB 1 ATGTCACACCCCGCGTAAACGCACTGCGTGCCTGCGCTATTGTAAAGAGATCCCTGCC 60

QY 26 LeuAspProGlnLeuAspTrpLeuLeuLeuAspSerMetThrLysArgPheGlu 45
DB 61 CTGGATCCGCAACTGCTCGACTGGCTGTCTGGAGGATTCATGACAAAACGTTTGA 120

QY 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
DB 121 CAGCAGGGAAACCGGTAAAGCGTACGATATCCCGAAGGGTTTGTCCAGCAGCAATGAA 180

QY 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
DB 181 ATCCCCGGAAGAACTGCCGCTGTCGCGAAAGAGTCTCGTTACTGTTACGTGAATTTTG 240

QY 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 105
DB 241 TTATGTCCGATGGTGAACCGTGGCTTCCCGGTGCTACCGTCTTCCTGTGTCAACGTTA 300

```

```

QY 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrIleuPhe 125
DB 301 AGCGGGCGGAGCTGGCGTTACAAAAATTGGTAAACCGCGTTAGGACGCTATCTGTTTC 360

QY 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyATGAspAlaGlyLeuTrpGly 145
DB 361 ACATCATCGACATTAAACCCGGGACTTTATTGAGATAGGCCGTGATGCCGGCTGTGGGGG 420

QY 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
DB 421 CGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTTACAGAACTGTTTTTACCG 480

QY 166 AlaSerProLeuTyr 170
DB 481 CGGTACCGTTGTAC 495

RESULT 10
ADV91618
ID ADV91618 standard; DNA; 498 BP.
XX
XX AC ADV91618;
XX
XX DT 10-MAR-2005 (first entry)
XX
XX DE Escherichia coli chorismate pyruvate lyase (CPL) DNA.
XX
XX KW Hydroquinone glucoside; arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside;
XX KW chorismate pyruvate lyase; CPL; 4-hydroxycinnamoyl-CoA hydratase/lyase;
XX KW HCHL; gene expression; hydroquinone; antiseptic; transgenic plant;
XX KW genetically engineered microorganism; antioxidant; antimicrobial agent;
XX KW anti-inflammatory agent; carcinogenic inhibitor; melanoma; cytostatic;
XX KW neoplasm; pHBA; para-hydroxybenzoic acid; p-hydroxybenzoic acid; Gene;
XX KW ds.
XX
XX OS Escherichia coli.
XX
XX FH Key Location/Qualifiers
XX CDS 1. .498
XX FT /*tag= a
XX FT /product= "Escherichia coli CPL protein"
XX
XX PN US2004261147-A1.
XX
XX PD 23-DEC-2004.
XX
XX PF 16-JUN-2003; 2003US-00462162.
XX
XX PR 16-JUN-2003; 2003US-00462162.
XX
XX PA (MEYE/) MEYER K.
XX PA (VIIT/) VIITANEN P V.
XX PA (FLIN/) FLINT D.
XX
XX PI Meyer K, Viitanen PV, Flint D;
XX
XX DR WPI; 2005-057232/06.
XX DR P-PSDB; ADV91619.
XX
XX PT Producing hydroquinone glucoside in a green plant comprises growing a
XX FT green plant having nucleic acid fragments and chorismate pyruvate lyase
XX FT expression cassette.
XX
XX PS Example 4; SEQ ID NO 29; 70pp; English.
XX
XX CC The present invention relates to methods and materials to produce
XX CC hydroquinone glucoside (arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside)
XX CC in genetically modified green plants and microorganisms. The method
XX CC relies upon transgenic plants or genetically modified microorganisms that
XX CC produce increased levels of the initial substrate para-hydroxybenzoic
XX CC acid (p-hydroxybenzoic acid; pHBA) in a biosynthetic pathway to produce
XX CC arbutin. Plants and microbes can be genetically engineered to produce
XX CC high levels of pHBA either by functional expression of the bacterial

```

CC protein chorismate pyruvate lyase (CPL) or by the expression of 4-  
CC hydroxycinnamoyl-CoA hydratase/lyase (HCHL) or a combination of both.  
CC Arbutin is useful as an antioxidant, antimicrobial agent, anti-  
CC inflammatory agent and inhibitor of carcinogenesis (melanoma). The  
CC present sequence is Escherichia coli chorismate pyruvate lyase (CPL) DNA.  
CC This sequence is found in the E. coli expression vector construct  
CC pET24a(+).

XX Sequence 498 BP; 114 A; 122 C; 141 G; 121 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 8.4e-89 Length: 498  
Score: 854.00 Matches: 165  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 96.3% Indels: 0  
DB: 14 Gaps: 0

US-10-718-311-16 (1-170) x ADV91618 (1-498)

QY 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 25  
DB 1 ATGTACACCCCGCTTACGCACTGCTGCGCTATGTAAAGAGATCCCTGCC 60  
QY 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45  
DB 61 CTGGATCCGCAACTGCTGACCTGGCTGTGCTGGAGGATTCATGACAAACGTTTGA 120  
QY 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAenGlu 65  
DB 121 CAGCAGGGAACAAACGTAAGCTGACGATCCGCAAGGGTGTTCGAGCAGATGAA 180  
QY 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeuLeu 85  
DB 181 ATCCCCGAGAACTGCGCTGTCGCGAAAGAGTCTCGTTACTGTTAGTGAATTTG 240  
QY 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 105  
DB 241 TTATGTGCGGATGGTGAACCGTGGCTTCCGCTGCTACCGTCTCTGTCACGTTA 300  
QY 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125  
DB 301 AGCGGCGCGAGCTGCGCTTACAAATAATGGTAAACCCGCTAGGACGCTATCTGTT 360  
QY 126 ThrSerSerThrLeuThrArgAspPheIleGluLeuGlyValArgAspAlaGlyLeuTrpGly 145  
DB 361 ACATCATCGACATTAAACCGGACCTTATTAGATAGGCCGTGATGCCGGCTGTGGGG 420  
QY 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165  
DB 421 CGACGTTCCCGCTCGATTAAAGCGTAAACCGCTGTGCTAACAGAACTGTTTTACCG 480  
QY 166 AlaSerProLeuTyr 170  
DB 481 GCGTCACCGTTGAC 495

#### RESULT 11

AAD06956

ID AAD06956 standard; DNA; 854 BP.

XX AAD06956;

AC AAD06956;

XX 06-AUG-2001 (first entry)

DE ubiC gene encoding chorismate pyruvate lyase.

XX p-hydroxybenzoic acid; pHb; aromatic pathway; carbon source; ubiC;  
KW chorismate pyruvate lyase; ppsA; phosphoenolpyruvate synthase;  
KW PEP synthase; tkkA; transketolase; aroC; chorismate synthase; aroA;  
KW 5-enolpyruvoylshikimate-3-phosphate synthase; BPS synthase; aroB;  
KW shikimate kinase II; aroB; 3-dehydroquinate synthase; DHQ synthase; aroG;  
KW 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase; DHAP synthase; ds.

OS Escherichia coli.  
XX US6210937-B1.  
XX 03-APR-2001.  
XX 22-APR-1998; 98US-00064693.  
XX 22-APR-1997; 97US-0044094P.  
XX (BECH-) BECHTEL BWXT IDAHO LLC.  
XX Ward TE, Watkins CS, Bulmer DK, Johnson BF, Amaratunga M;  
XX WPI; 2001-280857/29.  
XX Producing aromatic compounds, especially commercially acceptable levels  
XX of p-hydroxybenzoic acid, comprises developing genetically engineered  
XX bacteria that carry selected genes of the common aromatic pathway.

Claim 5; Col 31-32; 25pp; English.

CC The present invention relates to a method of producing p-hydroxybenzoic  
CC acid (pHB) which comprises transforming microorganisms with plasmids  
CC carrying selected genes of the common aromatic pathway and by conversion  
CC of carbon sources. The selected genes of the common aromatic pathway of  
CC the invention includes ubiC gene encoding chorismate pyruvate lyase, aroG  
CC structural gene encoding 3-deoxy-D-arabino-heptulosonate-7-phosphate  
CC (DHAP) isoenzyme synthase (phe), tkkA gene encoding transketolase, ppsA  
CC gene encoding phosphoenolpyruvate (PEP) synthase, aroC gene encoding  
CC chorismate synthase, aroI gene encoding shikimate kinase II, aroA gene  
CC encoding 5-enolpyruvoylshikimate-3-phosphate (BPS) synthase and aroB  
CC gene encoding 3-dehydroquinate (DHQ) synthase. The method is used for  
CC converting chorismate to a selected aromatic compound. The method is  
CC particularly used for the production of commercially acceptable levels of  
CC aromatic compounds, particularly p-hydroxybenzoic acid. The present  
CC sequence is ubiC gene which encodes chorismate pyruvate lyase that  
CC catalyses the production of chorismate from pHB

SQ Sequence 854 BP; 202 A; 200 C; 222 G; 230 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 1.78e-88 Length: 854  
Score: 854.00 Matches: 165  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 96.3% Indels: 0  
DB: 4 Gaps: 0

US-10-718-311-16 (1-170) x AAD06956 (1-854)

QY 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 25  
DB 321 ATGTACACCCCGCTTACGCACTGCTGCGCTATGTAAAGAGATCCCTGCC 380  
QY 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45  
DB 381 CTGGATCCGCAACTGCTGACCTGGCTGTGCTGGAGGATTCATGACAAACGTTTGA 440  
QY 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAenGlu 65  
DB 441 CAGCAGGGAACAAACGTAAGCTGACGATGATCCGCAAGGGTGTTCGAGCAGATGAA 500  
QY 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeuLeu 85  
DB 501 ATCCCCGAGAACTGCGCTGCTGCGAAAGAGTCTCGTTACTGTTAGTGAATTTG 560  
QY 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 105  
DB 561 TTATGTGCGGATGGTGAACCGTGGCTTCCGCGTCTGCTGCTGTTCAACGTTA 620  
QY 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125

Db 621 AGCGGCGGAGCTGCGCTTACAAAAATGGTAAACGCGCTTAGGACGCTATCTGTTTC 680  
 Qy 126 ThrSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145  
 Db 681 ACATCATCGACATTAAACCGGCGACTTTATTAGATAGGCCGCTGATGCCGCGCTGTGGGG 740  
 Qy 146 ArgArgSerArgLeuArgLeuSerGlyIlypProLeuLeuLeuThrGluLeuPheLeuPro 165  
 Db 741 CGAGTTTCCCGCTCGGATTAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTACC 800  
 Qy 166 AlaSerProLeuTyr 170  
 Db 801 GCGTCACCGTTGTAC 815  
 RESULT 12  
 AAD06957  
 ID AAD06957 standard; DNA; 6641 BP.  
 XX  
 AC AAD06957;  
 XX  
 DT 06-AUG-2001 (first entry)  
 XX  
 DE pME2 comprising ubiC, aroG, tktA and ppsA genes of the aromatic pathway.  
 XX  
 KW p-hydroxybenzoic acid; pHB; aromatic pathway; carbon source; ubiC;  
 KW chorismate pyruvate lyase; ppsA; phosphoenolpyruvate synthase;  
 KW PEP synthase; tktA; transketolase; aroC; chorismate synthase; aroA;  
 KW 5-enolpyruvoylshikimate-3-phosphate synthase; BPS synthase; aroL;  
 KW shikimate kinase II; aroB; 3-dehydroquininate synthase; DHQ synthase; aroG;  
 KW 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase; DHAP synthase;  
 KW pME2; ds.  
 XX  
 OS Escherichia coli.  
 XX  
 FN US6210937-B1.  
 XX  
 PD 03-APR-2001.  
 XX  
 PF 22-APR-1998; 98US-00064693.  
 XX  
 PR 22-APR-1997; 97US-0044094P.  
 XX  
 PA (BECH-) BECHTEL BWXT IDAHO LLC.  
 XX  
 PI Ward TE, Watkins CS, Bulmer DK, Johnson BF, Amaratunga M;  
 XX  
 DR WPI; 2001-280857/29.  
 XX  
 PT Producing aromatic compounds, especially commercially acceptable levels  
 PT of p-hydroxybenzoic acid, comprises developing genetically engineered  
 PT bacteria that carry selected genes of the common aromatic pathway.  
 XX  
 PS Claim 4; Col 31-40; 25pp; English.  
 XX  
 CC The present invention relates to a method of producing p-hydroxybenzoic  
 CC acid (pHB) which comprises transforming microorganisms with plasmids  
 CC carrying selected genes of the common aromatic pathway and by conversion  
 CC of carbon sources. The selected genes of the common aromatic pathway of  
 CC the invention includes ubiC gene encoding chorismate pyruvate lyase, aroG  
 CC structural gene encoding 3-deoxy-D-arabino-heptulosonate-7-phosphate  
 CC (DHAP) isoenzyme synthase (phe), tktA gene encoding transketolase, ppsA  
 CC gene encoding phosphoenolpyruvate (PEP) synthase, aroC gene encoding  
 CC chorismate synthase, aroL gene encoding shikimate kinase II, aroA gene  
 CC encoding 5-enolpyruvoylshikimate-3-phosphate (BPS) synthase and aroB  
 CC gene encoding 3-dehydroquininate (DHQ) synthase. The method is used for  
 CC converting chorismate to a selected aromatic compound. The method is  
 CC particularly used for the production of commercially acceptable levels of  
 CC aromatic compounds, particularly p-hydroxybenzoic acid. The present  
 CC sequence is pME2 plasmid which comprises ubiC, aroG, tktA and ppsA genes  
 CC involved in the aromatic pathway  
 XX  
 SQ Sequence 6641 BP; 1567 A; 1698 C; 1853 G; 1523 T; 0 U; 0 Other;

Alignment Scores: 3.08e-87 Length: 6641  
 Pred. No.: 854.00 Matches: 165  
 Score: 854.00 Conservative: 0  
 Percent Similarity: 100.0% Mismatches: 0  
 Best Local Similarity: 100.0% Indels: 0  
 Query Match: 96.3% Gaps: 0  
 DB: 4  
 US-10-718-311-16 (1-170) x AAD06957 (1-6641)  
 Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25  
 Db 321 ATGTCACACCCCGGCTTAACGCAACTGCGTGGCTGCGCTATTGTAAAGAGATCCCTGCC 380  
 Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 45  
 Db 381 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCCATGCACAAAACGTTTTTGA 440  
 Qy 46 GlnGlnGlyIlypThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65  
 Db 441 CAGCAGGGAACAAACGGTAAGCGTACGATGATCCGGAAGGGTTTGTGCGACAGAAAGAA 500  
 Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85  
 Db 501 ATCCCCGAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTTG 560  
 Qy 86 LeuCysAlaAspGlyGluProTrpTrpLeuAlaGlyArgThrValValProValSerThrLeu 105  
 Db 561 TTATGTCCGATGTTGAACCGTGGCTTCCGCGTCTGACCGCTGTTCTGTCAACGTTA 620  
 Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyIlypThrProLeuGlyArgTyrLeuPhe 125  
 Db 621 AGCGGCGGAGCTGCGCTTACAAAAATGGTAAACGCGCTTAGGACGCTATCTGTTTC 680  
 Qy 126 ThrSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145  
 Db 681 ACATCATCGACATTAAACCGGCGACTTTATTAGATAGGCCGCTGATGCCGCGCTGTGGGG 740  
 Qy 146 ArgArgSerArgLeuArgLeuSerGlyIlypProLeuLeuLeuThrGluLeuPheLeuPro 165  
 Db 741 CGAGTTTCCCGCTCGGATTAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTACC 800  
 Qy 166 AlaSerProLeuTyr 170  
 Db 801 GCGTCACCGTTGTAC 815  
 RESULT 13  
 ADV91643  
 ID ADV91643 standard; DNA; 1971 BP.  
 XX  
 AC ADV91643;  
 XX  
 DT 10-MAR-2005 (first entry)  
 XX  
 DE C. parapsilosis pHB 1-H- E. coli CPL chimeric DNA.  
 XX  
 KW Hydroquinone glucoside; arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside;  
 KW chorismate pyruvate lyase; CPL; 4-hydroxycinnamoyl-CoA hydratase/lyase;  
 KW HCHL; gene expression; hydroquinone; antiseptic; transgenic plant;  
 KW genetically engineered microorganism; antioxidant; antimicrobial agent;  
 KW anti-inflammatory agent; carcinogenic inhibitor; melanoma; cytostatic;  
 KW neoplasm; pHB 1-hydroxylase; pHB; para-hydroxybenzoic acid;  
 KW p-hydroxybenzoic acid; gene; ds.  
 XX  
 OS Candida parapsilosis.  
 OS Escherichia coli.  
 OS Chimeric.  
 XX  
 FH Location/Qualifiers  
 CDS 4..501  
 FT /\*tag= a  
 FT /product= "Escherichia coli CPL protein"  
 FT /transl\_except= (pos: 46..48, aa:Cys)



CC specifically the *Rhodobacter capsulatus* transformant MC9R/pRSFAC

Sequence 2000 BP; 444 A; 462 C; 538 G; 556 T; 0 U; 0 Other;;  
SQ  
XX

**Alignment Scores:**

Fragment Size (bp):	1,298-87	Length:	2000
Pred. No.:	851.00	Matches:	164
Score:	100.0%	Conservative:	1
Percent Similarity:	100.0%	Mismatches:	0
Best Local Similarity:	99.4%	Indels:	0
Query Match:	95.9%	Gaps:	0
DB:	2		

US-10-718-311-16 (1-170) x AAT29820 (1-2000)

Qy	6	MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla	25
Db	380	ATGTCTACACCCGCGTTACGCAACTGGCGTGCCTATTGTAAAGAGATCCCTGCGC	439
Qy	26	LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu	45
Db	440	CTGGATCCGCAACTGCTCGACTGCTGTGTGCTGAGGATTCATGACAAAACGTTTGGAA	499
Qy	46	GlnGlnGlySerValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu	65
Db	500	CAGCAGGGAAAAACGGTAAGCTGACGATATGGCGGAAGGGTTGTTCGAGCAGAAATGAA	559
Qy	66	IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu	85
Db	560	ATCCCGCAGAAGACTCGCGCTGCTGCCGAAGAGTCTCGTTACTGGTTACGTGAATTTTG	619
Qy	86	LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu	105
Db	620	TTATGTGCGCGATGGTGAACCGGTGGCTTCCCGGTGGTACCGTCTGTGTGTCACAGTTA	679
Qy	106	SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe	125
Db	680	AGCGGGCCGGAGCTGGCGTTACAAAAATTGGTAAACGCCGTTAGACCGCTATCTGTTC	739
Qy	126	ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly	145
Db	740	ACATCATCAGACATTAAACCGGACCTTTATTGAGATAGCGCGTGTGCGGCGCTGCGGG	799
Qy	146	ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro	165
Db	800	CGAGTTCCCGCGCTGGATTAGCCGTTAAACCGCTGTGTGCTTACAGAACTGTTTTACCG	859
Qy	166	AlaSerProLeuTyr	170
Db	860	GGGTACCGTTGTAC	874

**RESULT 15**

**AAS90335**

ID AAS90335 standard; cDNA; 1207 BP.

AC AAS90335;

XX

DT 13-FEB-2002 (first entry)

XX

DE DNA encoding novel human diagnostic protein #26139.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder;

OS Homo sapiens.

XX  
-----

PN WO200175067-

	7
	6
	5
	E
	C
	7
X	5
X	5

PD 11-OCT-2001.  
yy

XX  
PF 30-MAR-2001: 2001WO-US008631-

XX	(HYSE-) HYSEQ INC.
PA	
XX	
PI	Drmanac RT, Liu C, Tang YT;
XX	
DR	WPI; 2001-639362/73.
DR	P-PSDB; ABG26148.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity.

PS Claim 1: SEO ID NO 26139; 103pp; English.

The invention relates to isolated polynucleotide (I) and sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome mapping and in recombinant production of (II). The polynucleotide (I) and in diabetics as expressed sequence tags for identification of genes.

genes. (I) is useful in gene therapy activity of (II) or to treat disease

CC activity of (I) or (II) to treat disease states involving (I) or (II) as  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
xx  
SQ Sequence 1207 BP; 316 A; 265 C; 288 G; 338 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 2 62e-83 Length: 1207  
Score: 811.00 Matches: 163  
Percent Similarity: 98.2% Conservative: 1  
Best Local Similarity: 97.6% Mismatches: 1  
Query Match: 91.4% Indels: 2  
DB: 5 Gaps: 0

US-10-718-311-16 (1-170) x AAS90335 (1-1207)

Qy	6	MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysIysGluIleProAla	25
Db	652	ATGTCACACCCCGGTTAAACGCAACTCGCTGCGCTCGCTATTATTGAAGAGATCCCTGCC	711
Qy	26	LeuAspProGlnLeuLeuAspTrrPLeuLeuGluAspSerMetThrIysArgPheGlu	45
Db	712	CTGGATCCGCAACTGCTCGACTGCGCTGTGTCTGGAGGATCCATGACAAACCGTTTGAA	771
Qy	46	GlnGlnGlyIysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu	65
Db	772	CAGCAGGGAAAAACGGTAGCGGTGACGATATCCGGCAGGGGTTGTTCGAGCAGCAATGAA	831
Qy	66	IleProGluGlnLeuProLeuLeuProIysGluSerArgTyrTrpLeuArgGluIleLeu	85
Db	832	ATCCCCGAAGAACTGCGCGCTGCTGCCGAAGAGATCTCGTTACTGGTTACGTGAATTTTG	891
Qy	86	-LeuCysAlaAspGlyGluProTrpLeuAla-GlyArgThrValValProValSerThrL	105
Db	892	IGTATGTGCGGATGGCGAACCGCGGCTTGCCGGGTCTGTACCGTCGTTCCTGTGTCAACGT	951
Qy	105	euSerGlyProGluLeuAlaLeuGlnIysLeuGlyIysThrProLeuGlyArgTyrLeuP	125
Db	952	TAAAGCGGCGGAGCTGGCGCTTACAAAAATTGGTTAAAAACGCCGCTTAGGACGCTATCTGT	1011

QY 125 heThrSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpG 145  
 |||||  
 Db 1012 TCACATCATCGACATTAAACCGGACTTTATTGAGATAGGCCGTGATGCCGGCTGTGGG 1071  
 QY 145 lyArgArgSerArgLeuArgLeuSerClyIysProLeuLeuLeuThrGluLeuPheLeuP 165  
 |||||  
 Db 1072 GCGGAGTTCCCGCCTGCGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTAC 1131  
 QY 165 roAlaSerProLeuTyr 170  
 |||||  
 Db 1132 CGGCGTCACCGTTGTAC 1148

Search completed: May 30, 2006, 02:42:24  
 Job time : 753.985 secs



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 29, 2006, 20:59:50 ; Search time 725.015 Seconds

(without alignments)

2380.132 Million cell updates/sec

Title: US-10-718-311-4

Perfect score: 854

Sequence: 1 MSHPALQLRALRYCKEIPA.....RLSGKPLILITELFLPASPLY 165

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 5244920 seqs, 3486124231 residues

Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/abss/ABSSWEB\_epool/US10718311/runat\_26052006\_164914\_13428/app\_query.fasta\_1  
-DB=N Geneseq -OFMT=fasta -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -SPART=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs02p  
-USER=US10718311 -CGEN\_1\_1\_2295 -runat\_26052006\_164914\_13428 -NCFU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_8.\*

- 1: Geneseqn1980s.\*
- 2: Geneseqn1990s.\*
- 3: Geneseqn2000s.\*
- 4: Geneseqn2001as.\*
- 5: Geneseqn2001bs.\*
- 6: Geneseqn2002as.\*
- 7: Geneseqn2002bs.\*
- 8: Geneseqn2003as.\*
- 9: Geneseqn2003bs.\*
- 10: Geneseqn2003cs.\*
- 11: Geneseqn2003ds.\*
- 12: Geneseqn2004as.\*
- 13: Geneseqn2004bs.\*
- 14: Geneseqn2005s.\*
- 15: Geneseqn2006s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

#### ALIGNMENTS

RESULT 1  
AAQ92409  
ID AAQ92409 standard; DNA; 495 BP.  
XX  
AC AAQ92409;  
XX  
DT 11-JAN-1996 (first entry)  
XX  
DE E.coli chorismate-pyruvate lyase coding sequence.  
XX  
KW Chorismate-pyruvate lyase gene; ubiC gene; transgenic plant;  
KW tobacco mosaic virus resistance; antiviral activity;  
KW p-hydroxybenzoic acid; ss.  
XX  
OS Escherichia coli.  
XX  
FN DE4423022-C1.  
XX  
PD 24-MAY-1995.  
XX  
PF 30-JUN-1994; 94DE-04423022.  
XX  
PR 30-JUN-1994; 94DE-04423022.  
XX  
PA (HEID/) HEIDE L.  
XX  
PI Heide L, Siebert M, Severin K;

*have this spans noted*

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 29, 2006, 20:59:50 ; Search time 725.015 Seconds  
(without alignments)  
2380.132 Million cell updates/sec

Title: US-10-718-311-4

Perfect score: 854

Sequence: 1 MSHPALQTRALRYCKEIPA.....RLSGRPLLTLELPASPLY 165

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 5244920 seqs, 3486124231 residues

Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-O=/abs/ABSSWEB spool/US10718311/runat\_26052006\_164914\_13428/app\_query.fasta\_1  
-DB=N\_Geneseq -QFM=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p  
-USER=US10718311 @CN 1 1 2295 @runat\_26052006\_164914\_13428 -NCPU=6 -ICPU=3  
-NO\_WMAP -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq.8.\*

- 1: Geneseqn1980s.\*
- 2: Geneseqn1990s.\*
- 3: Geneseqn2000s.\*
- 4: Geneseqn2001as.\*
- 5: Geneseqn2001bs.\*
- 6: Geneseqn2002as.\*
- 7: Geneseqn2002bs.\*
- 8: Geneseqn2003as.\*
- 9: Geneseqn2003bs.\*
- 10: Geneseqn2003cs.\*
- 11: Geneseqn2003ds.\*
- 12: Geneseqn2004as.\*
- 13: Geneseqn2004bs.\*
- 14: Geneseqn2005s.\*
- 15: Geneseqn2006s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	854	100.0	495	2	AAQ92409 E.coli ch
2	854	100.0	495	6	ABA91837 Escherich
3	854	100.0	495	6	ABK47709 DNA seque

4	854	100.0	495	10	ADC25984	Adc25984 Escherich
5	854	100.0	498	14	ADV91618	Adv91618 Escherich
6	854	100.0	684	6	ABA91838	AbA91838 Chloropla
7	854	100.0	684	6	ABK47712	Abk47712 DNA seque
8	854	100.0	684	10	ADC25988	Adc25988 Tomato/Es
9	854	100.0	684	14	ADV91622	Adv91622 Tomato ru
10	854	100.0	854	4	AAD06956	Aad06956 ubiC gene
11	854	100.0	3452	14	ADV91636	Adv91636 C. paraps
12	854	100.0	6641	0	AAD06957	Aad06957 pHE2 comp
13	851	99.6	2000	2	AAT29820	Aat29820 E. coli u
14	843	98.7	1371	14	ADV91643	Adv91643 C. paraps
15	811	95.0	1207	5	AAS90335	Aas90335 DNA encod
16	790	92.5	1918	2	AAQ67671	Aaq67671 E.coli ub
17	673	78.8	645	11	ACH95439	Ach95439 Klebsiell
18	417	48.8	519	10	ACF70957	AcF70957 Photorhab
19	417	48.8	110000	10	ACF67367_40	Continuation (41 o
20	417	48.8	110000	10	ACF65388_07	Continuation (8 of
21	401	47.0	537	10	ADF02837	AdF02837 Bacterial
22	186	21.8	630	11	ABD08838	Abd08838 Pseudomon
23	174.5	20.4	519	12	ADL04203	Adl04203 DNA encod
24	174	20.4	100	8	ACD78861	AcD78861 E. coli K
25	171.5	20.1	269223	4	AAF28554	Aaf28554 Genomic f
26	167	19.6	100	8	ACD78860	AcD78860 E. coli K
27	165	19.3	100	8	ACD78859	AcD78859 E. coli K
28	157.5	18.4	540	9	ADA29056	Ada29056 DNA encod
29	146.5	17.2	1248	11	ABD08757	Abd08757 Pseudomon
30	130.5	15.3	609	12	ADN36906	Adn36906 X. albili
31	130.5	15.3	55839	12	ADN36893	Adn36893 X. albili
32	121	14.2	195	10	ADF02801	AdF02801 Bacterial
33	97.5	11.4	475	6	AAS97252	Aas97252 Neisseria
34	90	10.5	2127	4	ABA95459	Aba95459 Thermus t
35	90	10.5	9133	3	AAA81730	Aaa81730 N. mening
36	90	10.5	110000	3	AAA81490_07	Continuation (8 of
37	90	10.5	349880	3	AAF21608	Aaf21608 Neisseria
38	88	10.3	42811	12	ADQ97948	Adq97948 Human can
39	87	10.2	59589	11	ACN44724	Acn44724 Mouse gen
40	86.5	10.1	110000	10	ACF67367_43	Continuation (44 o
41	86.5	10.1	110000	10	ACF65388_04	Continuation (5 of
42	86	10.1	2892	15	AEF63232	Aef63232 Human UNC
43	85.5	10.0	522	10	ABZ33133	Abz33133 N. gonorr
44	85.5	10.0	2157	1	AAH70438	Aan70438 Streptoco
45	85	10.0	25785	4	AAK78465	Aak78465 Human imm

ALIGNMENTS

RESULT 1  
AAQ92409  
ID AAQ92409 standard; DNA; 495 BP.

AC AAQ92409;

DT 11-JAN-1996 (first entry)

DE E.coli chorismate-pyruvate lyase coding sequence.

KW Chorismate-pyruvate lyase gene; ubiC gene; transgenic plant;

KW tobacco mosaic virus resistance; antiviral activity;

KW p-hydroxybenzoic acid; ss.

OS Escherichia coli.

PN DE4423022-C1.

XX 24-MAY-1995.

PF 30-JUN-1994; 94DE-04423022.

PR 30-JUN-1994; 94DE-04423022.

PA (HEID/) HEIDE L.

PI Heide L, Siebert M, Severin K;

have this  
spms noted

XX WPI; 1995-186908/25.  
 DR P-PSDB; AAR74742.  
 XX Transgenic plants with increased content of resistance factor - contg.  
 PT bacterial DNA coding for enzyme catalysing prodn. of resistance factor,  
 PT e.g. antiviral p-hydroxy-benzoic acid.  
 XX Claim 3; Page 5; 7pp; German.  
 XX Transgenic plants containing a bacterial gene which encodes an enzyme  
 CC able to catalyse prodn. of an antiviral, bactericidal, fungicidal or  
 CC insecticidal factor are new. A preferred gene is ubiC from E.coli  
 CC (AAQ92409) which codes for chorismate-pyruvate lyase (AAR74742). The  
 CC lyase catalyses conversion of chorismate to p-hydroxybenzoic acid and  
 CC transgenic tobacco plants which express the active enzyme are resistant  
 CC to tobacco mosaic virus  
 XX Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 6.52e-89 Length: 495  
 Score: 854.00 Matches: 165  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 100.0% Indels: 0  
 DB: 2 Gaps: 0

US-10-718-311-4 (1-165) x AAQ92409 (1-495)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20  
 Db 1 ATGTCACACCCCGCGCTTAAACGCACTGCGTGCCTATTGTTAAAGAGATCCCTGCC 60  
 Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40  
 Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAACGTTTGAA 120  
 Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60  
 Db 121 CAGCAGGCGAAACCGTAAAGCGTGACGATGATCCGCGAAGGGTTTGTCCAGCAGATGAA 180  
 Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80  
 Db 181 ATCCCGCGAAGAACTGCCGCTGTCGCGAAAGAGTCTCGTTACTGTTACGTTAAATTTG 240  
 Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100  
 Db 241 TTATGTGCGGATGGTGAACCGTGGCTTCCCGTCCGTCGTTCCGTGTCACGTTA 300  
 Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrIleuPhe 120  
 Db 301 AGCGGGCGGAGCTGGCGCTTACAAAATTTGGTAAACCGCTTAGGACGCTATCTGTTC 360  
 Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140  
 Db 361 ACATCATCGATCAATTAACCCGGGACTTTATTGAGATAGCCGCTGATGCCGGCTGTGGGG 420  
 Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160  
 Db 421 CGACGTTCCCGCTGCGATTAAACCGGTAAACCGCTGTTGCTTAACAGAACTGTTTATCCG 480  
 Qy 161 AlaSerProLeuTyr 165  
 Db 481 CGGTACCGTGTATC 495

RESULT 2

ABA91837  
 ID ABA91837 standard; DNA; 495 BP.

XX ABA91837;

XX AC ABA91837;

XX TT 15-MAY-2002 (first entry)

XX Escherichia coli chorismate pyruvate lyase gene coding region.  
 DE  
 XX Chorismate pyruvate lyase; CPL; enzyme; p-hydroxybenzoic acid;  
 KW transgenic plant; ubiC; gene; ds.  
 XX Escherichia coli.  
 OS  
 XX Key Location/Qualifiers  
 FH 1. .495  
 CDS /\*tag= a  
 FT /partial  
 FT /product= "Chorismate\_pyruvate\_lyase"  
 FT /note= "the CDS does not include a stop codon"  
 XX WO200194607-A2.  
 XX 13-DEC-2001.  
 XX 22-MAY-2001; 2001WO-US016661.  
 XX 02-JUN-2000; 2000US-0209854P.  
 XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX Meyer K, Van Dyk DE, Viitanen PV;  
 XX WPI; 2002-226795/28.  
 XX P-PSDB; AAM50958.

Producing para-hydroxy benzoic acid in green plant. comprises expression

of unique expression cassette containing gene encoding chorismate  
 pyruvate lyase operably linked to specific chloroplast targeting  
 sequence.  
 Claim 4; Page 55; 60pp; English.  
 The present sequence is that of the coding region of the Escherichia coli  
 strain W3110 ubiC gene in expression construct pET24a. The gene encodes  
 chorismate pyruvate lyase (CPL, see AAM50958). It was obtained by PCR  
 amplification of strain W3110 genomic DNA using primers (see ABA91839-40)  
 based on the published E. coli ubiC gene. The CPL open reading frame is  
 used in an expression cassette designed for the high-level production of  
 p-hydroxybenzoic acid (pHBA) in green plants. The expression cassette  
 comprises the CPL coding sequence operably linked to a promoter capable  
 of driving protein expression in higher plants. The cassette also has a  
 sequence encoding a chloroplast transit peptide, its natural cleavage  
 site, and a small portion of a transit peptide donor protein fused to the  
 N-terminus of CPL. The chloroplast targeting sequence targets the foreign  
 protein to the chloroplast and aids in its uptake into the organelle. The  
 cleavage site is unique to the transit peptide, and cleavage of the  
 chimeric protein encoded by the cassette at this site releases a novel  
 polypeptide that has full enzyme activity, comprising the mature CPL  
 enzyme and a small portion of transit peptide donor (see AAM50961). A  
 plant comprising the CPL expression cassette is claimed, and may be  
 soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat,  
 barley, oats, sorghum, rice, Arabidopsis, sugarcane, canola,  
 millet, bean, pea, rye, flax or a forage grass. pHBA is a monomeric  
 component of liquid crystal polymers which have application in the  
 automotive, electrical and other industries

Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 6.52e-89 Length: 495  
 Score: 854.00 Matches: 165  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 100.0% Indels: 0  
 DB: 6 Gaps: 0

US-10-718-311-4 (1-165) x ABA91837 (1-495)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20  
Db 1 ATGTCAACCCCGCTTAAACCACTGCGTGCCTGCTATTTGTAAGAGATCCCTGCC 60  
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40  
Db 61 CTGGATCCGCNACTGCTCGACTGGCTGCTGCTGCGAGGATTCATGACAAACGTTTGAA 120  
QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60  
Db 121 CAGCAGGGAACCGTAAGCGTGACGATGATCCGCGAGGGTTTGTGAGCAGATGAA 180  
QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80  
Db 181 ATCCCCGAAGAACTGCGCTGCTGCGCGAAAGAGTCTCGTTACTGTGTACGTGAAATTTG 240  
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100  
Db 241 TTATGTGCCGATGGTGAACCGTGGCTTGGCGGTCGTACCGTCTGCTGTCAACGTTA 300  
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120  
Db 301 AGCGGCGCGAGTGGCTGTACAAAATTGGGTAAACCGCGTTAGAGCGCTATCTGTTTC 360  
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140  
Db 361 ACATCATCGACATTAACCCGGGACTTTATTGAGTAGGCGGTGATCCCGGCTGTGGGG 420  
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160  
Db 421 CGACGTTCCCGCTGCGATTAAGCGGTAAACCGCTGTGCTAACAGAACTGTTTTTACCG 480  
QY 161 AlaSerProLeuTyr 165  
Db 481 GCGTCACCGTTGTAC 495

RESULT 3

ID ABK47709 standard; DNA; 495 BP.  
XX AC ABK47709;  
XX 18-JUN-2002 (first entry)  
DE DNA sequence of ORF for CPL in pET24a E. coli expression construct.  
KW SNG1; sinapoylglucose accumulator 1; SMT; aromatic ester;  
KW sinapoylglucose:malate sinapoyltransferase; glycosylated aromatic acid;  
KW malate conjugated aromatic acid; polymer synthesis; ubiC; CPL; gene;  
KW carboxylic acid conjugated aromatic acid; chorisemate pyruvate lyase;  
KW pET24a; ds.  
OS Escherichia coli.  
XX FH Key Location/Qualifiers  
FT CDS 1. .495  
FT /tag= a  
FT /partial  
FT /product= "CPL"  
FT /note= "This sequence lacks a stop codon"  
XX WO200204653-A2.  
XX 17-JAN-2002.  
XX 05-JUL-2001; 2001WO-US021283.  
XX 07-JUL-2000; 2000US-0216615P.  
XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
XX Flint D, Meyer K, Viitanen PV;  
XX

WPI; 2002-303779/34.  
P-PSDB; AAU77941.

Producing aromatic acid conjugates, involves contacting glycosylated aromatic acid with malate, alpha-hydroxycarboxylic acid or alcohol in the presence of sinapoylglucose:malate sinapoyltransferase.

Example; Page 68; 72pp; English.

The present invention relates to the isolation of Arabidopsis thaliana gene (SNG1, sinapoylglucose accumulator 1) encoding sinapoylglucose:malate sinapoyltransferase (SMT). SMT catalyses the substitution of a glucose moiety on a glycosylated aromatic acid with a malate moiety to form a malate conjugated aromatic acid. The enzyme is useful for producing malate conjugated aromatic acids, carboxylic acid conjugated aromatic acids or aromatic esters. The malate conjugated aromatic acids are useful in the synthesis of various polymers. The present sequence representing the ORF (Open reading frame) for chorisemate pyruvate lyase (CPL) in pET24a Escherichia coli expression construct is described in the examples of the present invention

SQ Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6.52e-89 Length: 495  
Score: 854.00 Matches: 165  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 6 Gaps: 0

US-10-718-311-4 (1-165) x ABK47709 (1-495)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20  
Db 1 ATGTCAACCCCGCTTAAACCACTGCGTGCCTGCTATTTGTAAGAGATCCCTGCC 60  
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40  
Db 61 CTGGATCCGCNACTGCTCGACTGGCTGCTGCTGCGAGGATTCATGACAAACGTTTGAA 120  
QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60  
Db 121 CAGCAGGGAACCGTAAGCGTGACGATGATCCGCGAGGGTTTGTGAGCAGATGAA 180  
QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80  
Db 181 ATCCCCGAAGAACTGCGCTGCTGCGCGAAAGAGTCTCGTTACTGTGTACGTGAAATTTG 240  
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100  
Db 241 TTATGTGCCGATGGTGAACCGTGGCTTGGCGGTCGTACCGTCTGCTGTGCAACGTTA 300  
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120  
Db 301 AGCGGCGCGAGCTGGCTGTACAAAATTGGGTAAACCGCGCTTAGGACGCTATCTGTTTC 360  
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140  
Db 361 ACATCATCGACATTAACCCGGGACTTTATTGAGTAGGCGGTGATGCGCGGCTGTGGGG 420  
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160  
Db 421 CGACGTTCCCGCTGCGATTAAGCGGTAAACCGCTGTGCTAACAGAACTGTTTTTACCG 480  
QY 161 AlaSerProLeuTyr 165  
Db 481 GCGTCACCGTTGTAC 495  
RESULT 4  
ID ADC25984 standard; DNA; 495 BP.  
XX

AC ADC25984;  
 XX  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE Escherichia coli chorismate pyruvate lyase DNA.  
 XX  
 KW UDP-glucosyltransferase; PHBA; p-hydroxybenzoic acid ester glucoside;  
 KW liquid crystal polymer; LCP; methylparaben; preservative; food;  
 KW cosmetic industry; ds; gene; ubiC; CPL; chorismate pyruvate lyase.  
 XX  
 OS Escherichia coli.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..495  
 FT /tag= a  
 FT /partial  
 FT /product= "Escherichia coli chorismate pyruvate lyase  
 FT protein"  
 FT /note= "No stop codon"  
 XX  
 FN WO200306836-A2.  
 XX  
 PD 14-AUG-2003.  
 XX  
 PF 06-FEB-2003; 2003WO-US005863.  
 XX  
 PR 07-FEB-2002; 2002US-0355511P.  
 XX  
 PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX  
 PI Meyer K, Van Dyk DE, Viitanen PV;  
 XX  
 XX WPI; 2003-767259/72.  
 DR P-PSDB; ADC25985.  
 XX  
 PT New nucleic acid encoding UDP-glucosyltransferase, useful for preparing  
 PT cells that produce p-hydroxybenzoic acid glucose ester, also the new  
 PT enzymes.  
 XX  
 PS Example 9; SEQ ID NO 37; 161pp; English.  
 XX  
 CC The invention relates to a novel isolated nucleic acid that encodes a UDP  
 CC -glucosyltransferase. The method of the invention may be used to  
 CC transform microorganisms or green plant cells so that these produce a  
 CC higher level of high-volume chemicals or materials, such as PHBA (p-  
 CC hydroxybenzoic acid) ester glucoside which is an intermediate for PHBA, a  
 CC monomer for liquid crystal polymers (LCPs) and starting material for  
 CC methylparaben. Methylparaben is a preservative commonly used in the food  
 CC and cosmetic industries. The encoded enzymes may be used for in vitro  
 CC production of these compounds and for identifying similar enzymes by  
 CC sequence comparison. The current sequence is that of the Escherichia coli  
 CC chorismate pyruvate lyase (CPL; UbiC) DNA of the invention.  
 XX  
 SQ Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 6.52e-89 Length: 495  
 Score: 854.00 Matches: 165  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 100.0% Indels: 0  
 DB: 10 Gaps:  
 US-10-718-311-4 (1-165) x ADC25984 (1-495)  
 QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20  
 Db 1 ATGTCACACCCCGCGTAAACGCACTGGCTGGCTGGCTATTGTAAAGAGATCCCTGCC 60  
 QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40  
 Db 61 CTGGATCGGCAACTGCTCGACTGGCTGTGTCTGGAGGATTCATGACAAACGTTTTGAA 120

QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60  
 Db 121 CAGCAGGGAACGTAAGCGTGACGATGATCCGGAAGGTTTTCGACGAGATGAA 180  
 QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80  
 Db 181 ATCCCCGGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGCTTACGTGAAATTTTG 240  
 QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100  
 Db 241 TTATGTCCGATGGTGAACCGTGGCTTGGCCGCTGATCCGCTGCTCTGTCACCGTTA 300  
 QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120  
 Db 301 AGCGGCGGAGCTGGCGTTACAAAAATTGGTAAAAACGCCGTTAGGACGCTATCTGTTTC 360  
 QY 121 ThrSerThrLeuThrArgAspPheIleGluLeuGlyArgAspAlaGlyLeuTrpGly 140  
 Db 361 ACATCATCGACATTAAACCCGGGACTTTATTAGATAGCCCGTGATGCCGGGCTGTGGGG 420  
 QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160  
 Db 421 CGACGTTCCGCGCTGCGATTAAAGCGGTAAACCGCTGTGTGCTAACAGAACTGTTTTACCG 480  
 QY 161 AlaSerProLeuTyr 165  
 Db 481 CGGTACCGTTGTAC 495  
 RESULT 5  
 ADV91618  
 ID ADV91618 standard; DNA; 498 BP.  
 XX  
 AC ADV91618;  
 XX  
 DT 10-MAR-2005 (first entry)  
 DE Escherichia coli chorismate pyruvate lyase (CPL) DNA.  
 XX  
 KW Hydroquinone glucoside; arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside;  
 KW chorismate pyruvate lyase; CPL; 4-hydroxycinnamoyl-coA hydratase/lyase;  
 KW HCLH; gene expression; hydroquinone; antiseptic; transgenic plant;  
 KW genetically engineered microorganism; antioxidant; antimicrobial agent;  
 KW anti-inflammatory agent; carcinogenic inhibitor; melanoma; cytostatic;  
 KW neoplasm; PHBA; para-hydroxybenzoic acid; p-hydroxybenzoic acid; gene;  
 KW ds.  
 XX  
 OS Escherichia coli.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..498  
 FT /tag= a  
 FT /product= "Escherichia coli CPL protein"  
 FT  
 FT US2004261147-A1.  
 XX  
 XX 23-DEC-2004.  
 XX  
 XX 16-JUN-2003; 2003US-00462162.  
 XX  
 XX 16-JUN-2003; 2003US-00462162.  
 XX  
 PA (WEYE/) MEYER K.  
 PA (VIIT/) VIITANEN P V.  
 PA (FLIN/) FLINT D.  
 XX  
 PI Meyer K, Viitanen PV, Flint D;  
 XX  
 XX WPI; 2005-057232/06.  
 DR P-PSDB; ADV91619.  
 XX  
 PT Producing hydroquinone glucoside in a green plant comprises growing a  
 PT green plant having nucleic acid fragments and chorismate pyruvate lyase  
 PT expression cassette.



XX The present invention relates to methods and materials to produce  
PS hydroquinone glucoside (arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside)  
XX in genetically modified green plants and microorganisms. The method  
CC relies upon transgenic plants or genetically modified microorganisms that  
CC produce increased levels of the initial substrate para-hydroxybenzoic  
CC acid (p-hydroxybenzoic acid; pHBA) in a biosynthetic pathway to produce  
CC arbutin. Plants and microbes can be genetically engineered to produce  
CC high levels of pHBA either by functional expression of the bacterial  
CC protein chorismate pyruvate lyase (CPL) or by the expression of 4-  
CC hydroxycinnamoyl-CoA hydratase/lyase (HCHL) or a combination of both.  
CC Arbutin is useful as an antioxidant, antimicrobial agent, anti-  
CC inflammatory agent and inhibitor of carcinogenesis (melanoma). The  
CC present sequence is Escherichia coli chorismate pyruvate lyase (CPL) DNA.  
CC This sequence is found in the E. coli expression vector construct  
CC pET24a(+).

SQ Sequence 498 BP; 114 A; 122 C; 141 G; 121 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 6,57e-89 Length: 498  
Score: 854.00 Matches: 165  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 14 Gaps: 0

US-10-718-311-4 (1-165) x ADV91618 (1-498)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20  
Db 1 ATGTACACCCCGGTACCGACTCGTGGCTGCGTATTGTAAAGAGATCCCTGCC 60  
QY 21 LeuAppProGlnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 40  
Db 61 CTGGATCCGCACTGCTCGACTGGCTGTGGTGGAGATTCCATGACAAACGTTTGA 120  
QY 41 GlnGlnGlyThrValSerValThrMetIleArgGluGlyPheValGlnAenGlu 60  
Db 121 CAGCAGGAAACCGTAAAGCGTACGATGATCCGGAAGGGTGTGTCAGCAGATGAA 180  
QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80  
Db 181 ATCCCGAAGAACTGCCGTGCTCGCGAAGAGTCTCGTTACTGGTTAGTGAATTTG 240  
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100  
Db 241 TTATGTGCGCGATGTAACCGTGGCTGCGGTGCTGACCGTCTGCTGTCAACGTTA 300  
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120  
Db 301 AGCGGCGCGAGTGGCGTTACAAAATTTGGGTAAACCCCGTTAGGACGCTATCTGTT 360  
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyValArgAspAlaGlyLeuTrpGly 140  
Db 361 ACATCATCAGATTAAACCGGACCTTTATGAGATAGCGCGGTGTCGGGTGTTGGGG 420  
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuLeuLeuLeuLeuLeuPro 160  
Db 421 CGACGTCCTCCGCTCGATTAAAGCGGTAAACCGTGTGCTTAACAGAACTGTTTTTAC 480  
QY 161 AlaSerProLeuTyr 165  
Db 481 GCGTCACCGTTGAC 495

RESULT 6  
ABA91838  
ID ABA91838 standard; DNA; 684 BP.  
XX  
XX ABA91838;  
AC

DT 29-AUG-2003 (revised)  
DT 15-MAY-2002 (first entry)  
XX  
DE Chloroplast transit peptide-chorismate pyruvate lyase fusion gene.  
XX  
KW Chloroplast transit peptide; tomato; Rubisco; plant;  
KW ribulose-1,5-bisphosphate carboxylase; chorismate pyruvate lyase; CPL;  
KW enzyme; p-hydroxybenzoic acid; transgenic plant; gene; ds.  
XX  
OS Lycopersicon esculentum.  
OS Escherichia coli.  
OS Chimeric.  
XX  
FH Key Location/Qualifiers  
CDS 1..495  
FT /tag= a  
FT /product= "chloroplast-targeted CPL fusion"  
FT transit\_peptide 1..186  
FT /tag= b  
FT /note= "tomato Rubisco chloroplast targeting sequence"  
FT mat\_peptide 187..492  
FT /tag= c  
FT /note= "E. coli chorismate pyruvate lyase"  
FT  
XX WO200194607-A2.  
XX  
PD 13-DEC-2001.  
XX  
PF 22-MAY-2001; 2001WO-US016661.  
XX  
PR 02-JUN-2000; 2000US-0209854P.  
XX  
PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
XX  
PI Meyer K, Van Dyk DE, Viitanen PV;  
XX  
DR WPI; 2002-226795/28.  
DR P-PSDB; AAM50959.  
XX  
PT Producing para-hydroxy benzoic acid in green plant, comprises expression  
PT of unique expression cassette containing gene encoding chorismate  
PT pyruvate lyase operably linked to specific chloroplast targeting  
PT sequence.  
XX  
XX Claim 17; Page 57; 60pp; English.  
XX  
XX The present nucleotide sequence is that of an open reading frame encoding  
CC chloroplast-targeted chorismate pyruvate lyase (CPL) fusion protein TP-  
CC CPL (see ABA91841-42) was used to generate a DNA  
CC fragment corresponding to the transit peptide of tomato Rubisco small  
CC subunit and the first 4 amino acid residues of mature Rubisco. The DNA  
CC fragment was ligated into pET24a-CPL, which carries the Escherichia coli  
CC CPL open reading frame (see ABA91837). The TP-CPL construct is an example  
CC of expression cassettes of the invention that are designed for the high-  
CC level production of p-hydroxybenzoic acid (pHBA) in higher plants.  
CC Cleavage of TP-CPL in the chloroplast releases a novel polypeptide (see  
CC AAM50961) that has full enzyme activity, converting chorismate to pHBA. A  
CC claimed plant comprising a CPL expression cassette is selected from  
CC soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat,  
CC barley, oats, sorghum, rice, Arabidopsis, sugarcane, canola,  
CC millet, bean, pea, rye, flax or a forage grass. pHBA is a monomeric  
CC component of liquid crystal polymers which have application in the  
CC automotive, electrical and other industries. (Updated on 29-AUG-2003 to  
CC standardise OS field)  
XX  
SQ Sequence 684 BP; 160 A; 171 C; 179 G; 174 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 1,02e-88 Length: 684  
Score: 854.00 Matches: 165  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0

```
DB: 6 Gaps: 0
US-10-718-311-4 (1-165) x ABA91838 (1-684)
QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
DB 187 ATGTACACCCCGCGTTAACGCAACTGGTGGCTGCGCTATTGTAAAGAGATCCCTGCC 246
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
DB 247 CTGGATCCGCAACTGCTCGACTGGCTGTGGAGGATTCCATGACAAAACGTTTGA 306
QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
DB 307 CAGCAGGGAAACCGTAAGCGTGACGATGATCCGGAAGGTTTGTGAGCAGAGATGAA 366
QY 61 IleProGluGluLeuProLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
DB 367 ATCCCGGAGAACTGCGCTGTCGGAAGAGTCTCGTACTGTTACGTGAATTTTG 426
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
DB 427 TTATGTGCGGATGGTAACCGTGGCTTCCGCTCGTACCGTCTGTTCTGTCAACGTTA 486
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrIlePhe 120
DB 487 AGCGGGCCGAGCTGGCGTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 546
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
DB 547 ACATCATCGACATTAACCCCGGACTTTTATTGAGATAGCGCTGATCCCGCTGTCAACGTTA 606
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
DB 607 CGACGTTCCCGCTCGGATTAAGCGGTAACCGCTGTTGCTAAACAGAACTGTTTTTACCG 666
QY 161 AlaSerProLeuTyr 165
DB 667 GCGTCACCGTTGTAC 681
RESULT 7
ID ABK47712 standard; DNA; 684 BP.
XX
AC ABK47712;
XX
DT 18-JUN-2002 (first entry)
XX
DE DNA sequence of ORF for TP-CPL fusion protein.
XX
KW SNG1; sinapoylglucose accumulator 1; SMT; aromatic ester; plant;
KW sinapoylglucose:malate sinapoyltransferase; glycosylated aromatic acid;
KW malate conjugated aromatic acid; polymer synthesis; ubiC; TP-CPL; gene;
KW carboxylic acid conjugated aromatic acid; chorismate pyruvate lyase;
KW tomato Rubisco small subunit precursor; rbcS2; mutant; ds.
XX
OS Escherichia coli.
OS Lycopersicon esculentum.
OS Synthetic.
OS Chimeric.
XX
FH Key Location/Qualifiers
CDS 1..684
FT /*tag= a
FT /product= "TP-CPL fusion protein"
XX
XX
PN WO200204653-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021283.
XX
PR 07-JUL-2000; 2000US-0216615P.
XX
```

(DUPO ) DU PONT DE NEMOURS & CO E I.  
Flint D, Meyer K, Viitanen PV;  
WPI: 2002-303779/34.  
P-PSDB; RAU77942.

Producing aromatic acid conjugates, involves contacting glycosylated aromatic acid with malate, alpha-hydroxycarboxylic acid or alcohol in the presence of sinapoylglucose:malate sinapoyltransferase.

Example; Page 70; 72pp; English.

The present invention relates to the isolation of Arabidopsis thaliana gene (SNG1, sinapoylglucose accumulator 1) encoding sinapoylglucose:malate sinapoyltransferase (SMT). SMT catalyses the substitution of a glucose moiety on a glycosylated aromatic acid with a malate moiety to form a malate conjugated aromatic acid. The enzyme is useful for producing malate conjugated aromatic acids, carboxylic acid conjugated aromatic acids or aromatic esters. The malate conjugated aromatic acids are useful in the synthesis of various polymers. The present sequence representing the ORF (open reading frame) for TP-CPL (tomato Rubisco small subunit precursor for rbcS2-Escherichia coli chorismate pyruvate lyase (CPL)) fusion protein is described in the examples of the present invention

Sequence 684 BP; 160 A; 171 C; 179 G; 174 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 1,02e-88 Length: 684  
Score: 854.00 Matches: 165  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: Gaps: 0

US-10-718-311-4 (1-165) x ABK47712 (1-684)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20  
DB 187 ATGTACACCCCGCGTTAACGCAACTGGTGGCTGCGCTATTGTAAAGAGATCCCTGCC 246  
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40  
DB 247 CTGGATCCGCAACTGCTCGACTGGCTGTGGAGGATTCCATGACAAAACGTTTGA 306  
QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60  
DB 307 CAGCAGGGAAACCGTAAGCGTGACGATGATCCGGAAGGTTTGTGAGCAGAGATGAA 366  
QY 61 IleProGluGluLeuProLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80  
DB 367 ATCCCGGAGAACTGCGCTGTCGGAAGAGTCTCGTACTGTTACGTGAATTTTG 426  
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100  
DB 427 TTATGTGCGGATGGTAACCGTGGCTTCCGCTCGTACCGTCTGTTCTGTCAACGTTA 486  
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrIlePhe 120  
DB 487 AGCGGGCCGAGCTGGCGTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 546  
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140  
DB 547 ACATCATCGACATTAACCCCGGACTTTTATTGAGATAGCGCTGATCCCGCTGTCAACGTTA 606  
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160  
DB 607 CGACGTTCCCGCTCGGATTAAGCGGTAACCGCTGTTGCTAAACAGAACTGTTTTTACCG 666  
QY 161 AlaSerProLeuTyr 165  
DB 667 GCGTCACCGTTGTAC 681

Db 667 GCGTCACCGTTGTAC 681

RESULT 8

ADC25988

ID ADC25988 standard; DNA; 684 BP.

XX

AC ADC25988;

XX

DT 18-DEC-2003 (first entry)

XX

DE Tomato/Escherichia coli chimeric TP-CPL DNA.

XX

KW UDP-glucosyltransferase; PHBA; p-hydroxybenzoic acid ester glucoside;

KW liquid crystal polymer; LCP; methylparaben; preservative; food;

KW cosmetic industry; ds; gene; tomato; TP-CPL; chimeric;

KW chorismate pyruvate lyase; plant.

XX

OS Chimeric.

OS Lycopersicon esculentum.

OS Escherichia coli.

XX

Key Location/Qualifiers

CDS 1..684

FT /tag= a

FT /product= "Tomato/Escherichia coli chimeric TP-CPL

FT protein"

XX

PN WO2003066836-A2.

XX

PD 14-AUG-2003.

XX

PF 06-FEB-2003; 2003WO-US005863.

XX

PR 07-FEB-2002; 2002US-0355511P.

XX

PA (DUPO ) DU PONT DE NEMOURS & CO E I.

XX

PI Meyer K, Van Dyk DE, Viitanen PV;

XX

XX WPI; 2003-767259/72.

DR P-PSDB; ADC25989.

DR

XX

PT New nucleic acid encoding UDP-glucosyltransferase, useful for preparing

PT cells that produce p-hydroxybenzoic acid glucose ester, also the new

PT enzymes.

XX

PS Example 9; SEQ ID NO 41; 161pp; English.

XX

CC The invention relates to a novel isolated nucleic acid that encodes a UDP

CC -glucosyltransferase. The method of the invention may be used to

CC transform microorganisms or green plant cells so that these produce a

CC higher level of high-volume chemicals or materials, such as PHBA (p-

CC hydroxybenzoic acid) ester glucoside which is an intermediate for PHBA, a

CC monomer for liquid crystal polymers (LCPs) and starting material for

CC methylparaben. Methylparaben is a preservative commonly used in the food

CC and cosmetic industries. The encoded enzymes may be used for in vitro

CC production of these compounds and for identifying similar enzymes by

CC sequence comparison. The current sequence is that of the

CC tomato/Escherichia coli chimeric TP-CPL (chorismate pyruvate lyase) DNA

CC of the invention.

XX

SQ Sequence 684 BP; 160 A; 171 C; 179 G; 174 T; 0 U; 0 Other;

XX

Alignment Scores:

Pred. No.: 1.02e-88 Length: 684

Score: 854.00 Matches: 165

Percent Similarity: 100.0% Conservative: 0

Best Local Similarity: 100.0% Mismatches: 0

Query Match: 100.0% Indels: 0

DB: 10 Gaps: 0

US-10-718-311-4 (1-165) x ADC25988 (1-684)

1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20

187 ATGTACACCCCGGTTAACGCAACTGCGTGGCTATTGTAAAGAGATCCCTGCC 246

21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40

247 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGATTCCATGACAAAACGTTTGA 306

41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAsnGlu 60

307 CAGCAGGGGAAACCGTAAGCGTGACGATCCGCGAAGGGTTTCTCGAGCAAGATGA 366

61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80

367 ATCCCCGGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTG 426

81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100

427 TTATGTGCGGATGGTGAACCGTGCGTTCGCGTTCGTACCGTCTCTGTGTCAACGTTA 486

101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120

487 ACGGGCGGAGCTGCGTTACAAAATTGGGTAAACCCCGTAGGACGCTATCTGTTTC 546

121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140

547 ACATCATCGACATTAAACCGGACTTTATTGAGATAGCGCGTGTGCGGGCTGTGGGG 606

141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160

607 CGACGTTCCCGCTCGGATTAGCGGTAAACCGCTGTTCTGAACAGAACTGTTTTTACCG 666

161 AlaSerProLeuTyr 165

667 GCGTCACCGTTGTAC 681

RESULT 9

ADV91622

ID ADV91622 standard; DNA; 684 BP.

XX

AC ADV91622;

XX

DT 10-MAR-2005 (first entry)

XX

DE Tomato rubisco small subunit precursor TP-E. coli CPL chimeric DNA.

XX

KW Hydroquinone glucoside; arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside;

KW chorismate pyruvate lyase; CPL; 4-hydroxycinnamoyl-CoA hydratase/lyase;

KW HCHL; gene expression; hydroquinone; antiseptic; transgenic plant;

KW genetically engineered microorganism; antioxidant; antimicrobial agent;

KW anti-inflammatory agent; carcinogenic inhibitor; melanoma; cytostatic;

KW neoplasm; PHBA; para-hydroxybenzoic acid; p-hydroxybenzoic acid;

KW rubisco small subunit precursor; transit peptide;

KW chloroplast transit peptide; TP; chloroplast; gene; ds.

XX

OS Lycopersicon esculentum.

OS Escherichia coli.

OS Chimeric.

XX

Key Location/Qualifiers

CDS 1..684

FT /tag= a

FT /product= "Tomato rubisco small subunit precursor TP-E.

FT coli CPL fusion protein"

XX

US2004261147-A1.

23-DEC-2004.

16-JUN-2003; 2003US-00462162.

16-JUN-2003; 2003US-00462162.

```

PA (MEYE/) MEYER K.
PA (VIIT/) VIITANEN P V.
PA (FLIN/) FLINT D.
XX
XX Meyer K, Viitanen PV, Flint D;
XX
XX WPI; 2005-057232/06.
XX P-PSDB; ADV91623.
XX
XX Producing hydroquinone glucoside in a green plant comprises growing a
XX green plant having nucleic acid fragments and chorismate pyruvate lyase
XX expression cassette.
XX
XX Example 4; SEQ ID NO 33; 70pp; English.
XX
XX The present invention relates to methods and materials to produce
XX hydroquinone glucoside (arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside)
XX in genetically modified green plants and microorganisms. The method
XX relies upon transgenic plants or genetically modified microorganisms that
XX produce increased levels of the initial substrate para-hydroxybenzoic
XX acid (p-hydroxybenzoic acid; PHBA) in a biosynthetic pathway to produce
XX arbutin. Plants and microbes can be genetically engineered to produce
XX high levels of PHBA either by functional expression of the bacterial
XX protein chorismate pyruvate lyase (CPL) or by the expression of 4-
XX hydroxycinnamoyl-CoA hydratase/lyase (HCHL) or a combination of both.
XX Arbutin is useful as an antioxidant, antimicrobial agent, anti-
XX inflammatory agent and inhibitor of carcinogenesis (melanoma). The
XX present sequence is tomato rubisco small subunit precursor (rbcS2).
XX transit peptide (chloroplast transit peptide; TP)-Escherichia coli CPL
XX chimeric DNA. This sequence is present in the E. coli expression vector
XX construct pET24a-TP-CPL.
XX
XX SQ Sequence 684 BP; 160 A; 171 C; 179 G; 174 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.02e-88 Length: 684
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 14 Gaps: 0

US-10-718-311-4 (1-165) x ADV91622 (1-684)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 187 ATGTACACCCCGCGCTTAACGCAACTGCGTGGCTGCTATTGTAAAGAGATCCCTGCC 246
Qy 21 LeuAspProGlnLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 247 CTGGATCCGCAACTGCTCGACTGGCTGTGTGGAGGATTCATGACAAAACGTTTGA 306
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 307 CAGCAGGGAANAACGGTAAGCGTGACGATGATCCGGAAGGTTTGTGACACAGATGAA 366
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 367 ATCCCGCAAGAACTGCCGCTGTCGCAAAAGAGTCTCGTTACTGCTTACGTGAAATTTG 426
Qy 81 LeuCybAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 427 TTATGTCCGATGGTGAACCGTGGCTTCCCGGTGCTACCGTCTGTGTCTCAACGTTA 486
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 487 AGCGGCGCGAGCTGGCGTTACAAAATTGGTAAACGCCGTTAGGACGCTATCTGTT 546
Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 547 ACATCATCGACATTAAACCGGACTTTATTAGATAGCCGCTGATGCCGGGCTGTGGGG 606
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuLeuThrGluLeuPheLeuPro 160

```

---

```

Db 607 CGACGTTCCCGCCTGCAGATTAAAGCGTAACCGCTGTGCTTAACAGAACTGTTTTACCG 666
Qy 161 AlaSerProLeuTyr 165
Db 667 GCGTCACCGTTGTAC 681

RESULT 10
AAD06956
ID AAD06956 standard; DNA; 854 BP.
XX
XX AC AAD06956;
XX
XX DT 06-AUG-2001 (first entry)
XX
XX ubiC gene encoding chorismate pyruvate lyase.
XX
XX p-hydroxybenzoic acid; PHB; aromatic pathway; carbon source; ubiC;
XX chorismate pyruvate lyase; ppsA; phosphoenolpyruvate synthase;
XX PEP synthase; tktA; transketolase; aroC; chorismate synthase; aroA;
XX 5-enolpyruvoylshikimate-3-phosphate synthase; EPSP synthase; aroL;
XX shikimate kinase II; aroB; 3-dehydroquinate synthase; DHQ synthase; aroG;
XX 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase; DHAP synthase; ds.
XX Escherichia coli.
XX
XX US6210937-B1.
XX
XX 03-APR-2001.
XX
XX 22-APR-1998; 98US-00064693.
XX
XX 22-APR-1997; 97US-0044094P.
XX
XX (BECH-) BECHTEL BWXT IDAHO LLC.
XX
XX Ward TE, Watkins CS, Bulmer DK, Johnson BF, Amaratunga M;
XX WPI; 2001-280857/29.
XX
XX Producing aromatic compounds, especially commercially acceptable levels
XX of p-hydroxybenzoic acid, comprises developing genetically engineered
XX bacteria that carry selected genes of the common aromatic pathway.
XX
XX Claim 5; Col 31-32; 25pp; English.
XX
XX The present invention relates to a method of producing p-hydroxybenzoic
XX acid (pHB) which comprises transforming microorganisms with plasmids
XX carrying selected genes of the common aromatic pathway and by conversion
XX of carbon sources. The selected genes of the common aromatic pathway of
XX the invention includes ubiC gene encoding chorismate pyruvate lyase, aroG
XX structural gene encoding 3-deoxy-D-arabino-heptulosonate-7-phosphate
XX (DHAP) isoenzyme synthase (phe), tktA gene encoding transketolase, ppsA
XX gene encoding phosphoenolpyruvate (PEP) synthase, aroC gene encoding
XX chorismate synthase, aroL gene encoding shikimate kinase II, aroA gene
XX encoding 5-enolpyruvoylshikimate-3-phosphate synthase (EPSP) and aroB
XX gene encoding 3-dehydroquinate (DHQ) synthase. The method is used for
XX converting chorismate to a selected aromatic compound. The method is
XX particularly used for the production of commercially acceptable levels of
XX aromatic compounds, particularly p-hydroxybenzoic acid. The present
XX sequence is ubiC gene which encodes chorismate pyruvate lyase that
XX catalyzes the production of chorismate from PHB
XX
XX SQ Sequence 854 BP; 202 A; 200 C; 222 G; 230 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.39e-88 Length: 854
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 4 Gaps: 0

```

```
US-10-718-311-4 (1-165) x AAD06956 (1-854)
QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 321 ATGTCACACCCCGGTAAAGCAACTGCGTGGCTATTTGTAAGAGATCCCTGCC 380
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 381 CTGGATCCGCAACTGCTCGACTGGCTTGGTCTGGAGGATTCATGACAAAACGTTTGA 440
QY 41 GlnGlnGlyThrValSerValThrMetIleArgGluGlyPheValGlnGlnAenGlu 60
Db 441 CAGCAGGGAACGATGACGATGATCCGGAAGGGTTTGTCCGACGAGATGAA 500
QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 501 ATCCCGAAGAACTGCGGTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTG 560
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 561 TTATGTGCCGATGGTGAACCGTGGCTTGGCGGTGCTACCGTCTCTCTGTCAACGTTA 620
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 621 AGCGGCGCGAGCTGGCGTTACAAAATTTGGGTAAACCGCGTTAGGACGCTATCTGTT 680
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 681 ACATCATCATCATTAACCGGAGCTTTATTTAGATAGGCGGTGATCCGGCTGTGGGG 740
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 741 CGACGTCTCCGCGCTCGGATTAAGCGGTAAACCGCTGTGTGTAACAGAACTGTTTTTAC 800
QY 161 AlaSerProLeuTyr 165
Db 801 GCGTCACCGTTGTAC 815

RESULT 11
ADV91636
ID ADV91636 standard; DNA; 3452 BP.
XX
AC ADV91636;
XX
DT 10-MAR-2005 (first entry)
XX
DE C. parapsilosis pHB 1-H- E. coli CPL- A. thaliana UGT72B1 chimeric DNA.
XX
KW Hydroquinone glucoside; arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside;
KW chorismate pyruvate lyase; CPL; 4-hydroxycinnamoyl-CoA hydratase/lyase;
KW HCHL; gene expression; hydroquinone; antiseptic; transgenic plant;
KW genetically engineered microorganism; antioxidant; antimicrobial agent;
KW anti-inflammatory agent; carcinogenic inhibitor; melanoma; cytostatic;
KW neoplasm; pHBA 1-hydroxylase; pHBA 1-H; pHBA; para-hydroxybenzoic acid;
KW p-hydroxybenzoic acid; UDP-glucosyltransferase; UGT72B1; gene; ds.
XX
OS Arabidopsis thaliana.
OS Candida parapsilosis.
OS Escherichia coli.
OS Chimeric.
XX
FH Key
FT CDS
FT Location/Qualifiers
FT 4..501
FT /tag= a
FT /product= "Escherichia coli CPL protein"
FT 541..1980
FT /tag= b
FT /product= "Candida parapsilosis pHBA 1-H protein"
FT /transl_except= (pos: 559..561, aa:Trp)
FT 2004..3446
FT /tag= c
FT /product= "Arabidopsis thaliana UGT72B1 protein"
XX

US-10-718-311-4 (1-165) x ADV91636 (1-3452)
QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 4 ATGTCACACCCCGGTAAAGCAACTGCGTGGCTATTTGTAAGAGATCCCTGCC 63
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 64 CTGGATCCGCAACTGCTCGACTGGCTTGGTCTGGAGGATTCATGACAAAACGTTTGA 123
QY 41 GlnGlnGlyThrValSerValThrMetIleArgGluGlyPheValGlnGlnAenGlu 60
Db 124 CAGCAGGGAACGATGACGATGATCCGGAAGGGTTTGTCCGACGAGATGAA 183
QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 184 ATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTG 243
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 244 TTATGTGCCGATGGTGAACCGGTGGCTTGGCGGTGCTACCGTCTCTCTGTCAACGTTA 303
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120

Alignment Scores:
Pred. No.: 9.61e-88 Length: 3452
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 14 Gaps: 0
```

Db 304 AGCGGGCGGAGCTGGCTTACAAAAATTGGTAAACCGCGTTAGGACGCTATCTGTTC 363  
 Qy 121 ThrSerThrLeuThrArgAspPheleGluileGlyArgAspAlaGlyLeuTrpGly 140  
 Db 364 ACATCATCGACATTACCCGGGACTTTATTGAGATAGGCGGTGATGCGGGCTGTGGGG 423  
 Qy 141 ArgArgSerArgLeuArgLeuSerGlyysProLeuLeuLeuThrGluLeuPheLeuPro 160  
 Db 424 CGAGCTTCCCGCTCGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTACCG 483  
 Qy 161 AlaSerProLeuTyr 165  
 Db 484 CGGTACCGTGTATC 498  
 RESULT 12  
 AAD06957  
 ID AAD06957 standard; DNA; 6641 BP.  
 XX  
 AC AAD06957;  
 XX  
 DT 06-AUG-2001 (first entry)  
 XX  
 DE pME2 comprising ubiC, aroG, tktA and ppsA genes of the aromatic pathway.  
 XX  
 KW p-hydroxybenzoic acid; pHb; aromatic pathway; carbon source; ubiC;  
 KW chorismate pyruvate lyase; ppsA; phosphoenolpyruvate synthase;  
 KW PEP synthase; tktA; transketolase; aroC; chorismate synthase; aroA;  
 KW 5-enolpyruvoylshikimate-3-phosphate synthase; PEP synthase; aroL;  
 KW shikimate kinase II; aroB; 3-dehydroquinate synthase; DHQ synthase; aroG;  
 KW 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase; DHAP synthase;  
 KW pME2; ds.  
 XX  
 OS Escherichia coli.  
 XX  
 PN US6210937-B1.  
 XX  
 PD 03-APR-2001.  
 XX  
 PF 22-APR-1998; 98US-00064693.  
 XX  
 PR 22-APR-1997; 97US-0044094P.  
 XX  
 PA (BECH-) BECHTEL BWXT IDAHO LLC.  
 XX  
 PI Ward TE, Watkins CS, Bulmer DK, Johnson BF, Amaratunga M;  
 XX  
 DR WPI; 2001-280857/29.  
 XX  
 PT Producing aromatic compounds, especially commercially acceptable levels  
 PT of p-hydroxybenzoic acid, comprises developing genetically engineered  
 PT bacteria that carry selected genes of the common aromatic pathway.  
 XX  
 PS Claim 4; Col 31-40; 25pp; English.  
 XX  
 CC The present invention relates to a method of producing p-hydroxybenzoic  
 CC acid (pHB) which comprises transforming microorganisms with plasmids  
 CC carrying selected genes of the common aromatic pathway and by conversion  
 CC of carbon sources. The selected genes of the common aromatic pathway of  
 CC the invention includes ubiC gene encoding chorismate pyruvate lyase, aroG  
 CC structural gene encoding 3-deoxy-D-arabino-heptulosonate-7-phosphate  
 CC (DHAP) isoenzyme synthase (phe), tktA gene encoding transketolase, ppsA  
 CC gene encoding phosphoenolpyruvate (PEP) synthase, aroC gene encoding  
 CC chorismate synthase, aroL gene encoding shikimate kinase II, aroA gene  
 CC encoding 5-enolpyruvoylshikimate-3-phosphate (PEP) synthase and aroB  
 CC gene encoding 3-dehydroquinate (DHQ) synthase. The method is used for  
 CC converting chorismate to a selected aromatic compound. The method is  
 CC particularly used for the production of commercially acceptable levels of  
 CC aromatic compounds, particularly p-hydroxybenzoic acid. The present  
 CC sequence is pME2 plasmid which comprises ubiC, aroG, tktA and ppsA genes  
 CC involved in the aromatic pathway  
 XX  
 SQ Sequence 6641 BP; 1567 A; 1698 C; 1853 G; 1523 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 2.38e-87 Length: 6641  
 Score: 854.00 Matches: 165  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 100.0% Indels: 0  
 DB: 4 Gaps: 0  
 US-10-718-311-4 (1-165) x AAD06957 (1-6641)  
 Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluileProAla 20  
 Db 321 ATGTACACCCCGGTTAAACGCACTGCTGGCTGGCTATTGTAAAGAGATCCCTGCC 380  
 Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40  
 Db 381 CTGGATCCGCAACTGCTCGACTGGCTGTTCTGGAGGATTCATGACAAAACGTTTTGAA 440  
 Qy 41 GlnGlnGlyysThrValSerValThrMetileargGluGlyPheValGlnGlnGlu 60  
 Db 441 CAGCAGGGAAACCGTAAGCGTGACGATGATCCGGAAGGGTTTGTCCGAGCAGAAATGAA 500  
 Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluileLeu 80  
 Db 501 ATCCCCGAGAACTGCGCTGCTCCGNAAGAGTCTCGTTACTGGTTACGTGAATTTTG 560  
 Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100  
 Db 561 TTATGTCCGATGTTGAACCGTGGCTTCCCGCTGTTACCGTCTGTGTGTCAACGTTA 620  
 Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyysThrProLeuGlyArgTyrLeuPhe 120  
 Db 621 AGCGGGCGGAGCTGGCGCTTACAAAATTGGTAAAAACGCCGTTAGGACGCTATCTGTT 680  
 Qy 121 ThrSerSerThrLeuThrArgAspPheleGluileGlyArgAspAlaGlyLeuTrpGly 140  
 Db 681 ACATCATCGACATTAAACCGGAGCTTTATTGAGATAGGCGGTGATGCCGGCTGTGGGG 740  
 Qy 141 ArgArgSerArgLeuArgLeuSerGlyysProLeuLeuLeuThrGluLeuPheLeuPro 160  
 Db 741 CGAGCTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTACCG 800  
 Qy 161 AlaSerProLeuTyr 165  
 Db 801 GCGTCACCGTGTATC 815  
 RESULT 13  
 AAT29820  
 ID AAT29820 standard; DNA; 2000 BP.  
 XX  
 AC AAT29820;  
 XX  
 DT 17-JAN-1997 (first entry)  
 XX  
 DE E. coli ubiquinone biosynthetic enzyme genes ubiC and ubiA.  
 XX  
 KW Ubiquinone-10; biosynthetic enzyme; ubiC; ubiA; gene;  
 KW photosynthetic bacteria; recombinant production; plasmid pRSPAC;  
 KW glutamate synthase; gene promoter; Rhodobacter capsulatus; transformant;  
 KW MCS9/pRSPAC; ss.  
 XX  
 OS Escherichia coli.  
 XX  
 Key Location/Qualifiers  
 CDS 380..877  
 FT /\*tag= a  
 FT /label= ubiC  
 CDS 890..1762  
 FT /\*tag= b  
 FT /label= ubiA  
 FT 1868..1887  
 FT /\*tag= c  
 FT repeat\_unit  
 FT /\*tag= inverted  
 FT /\*tag= c

```

FT repeat_unit 1892..1911
FT /*tag= d
FT /*rpt_type= inverted
PN JP08107789-A.
XX
XX
XX 30-APR-1996.
XX
XX 13-OCT-1994; 94JP-00273071.
XX
XX 13-OCT-1994; 94JP-00273071.
XX
XX (ALPH-) ALPHA SHOKUHIN KK.
XX
XX WPI; 1996-262599/27.
XX P-PSDB; AAR97746, AAR97747.
XX
XX Genes coding for ubiquinone biosynthetic enzymes - useful for prodn. of
XX ubiquinone-10 by transformed photosynthetic bacteria.
XX
XX Claim 1; Fig 2; lipp; Japanese.
XX
XX The present sequence comprises the ubiquinone biosynthetic enzyme genes
XX ubiC and ubiA, which were cloned from the chromosomal DNA of E. coli
XX (Kohara map phage DNA bank IF8 (634); Cell 50, 495-508 (1987)). A large
XX amt. of ubiquinone-10 can be produced by culturing photosynthetic
XX bacteria transformed with the novel plasmid pRSPAC, which contains the
XX ubiC and ubiA genes and the glutamate synthase gene promoter.
XX Specifically the Rhodobacter capsulatus transformant MC9R/pRSPAC
XX
XX Sequence 2000 BP; 444 A; 462 C; 538 G; 556 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 1e-87 Length: 2000
XX Score: 851.00 Matches: 164
XX Percent Similarity: 100.0% Conservative: 1
XX Best Local Similarity: 99.4% Mismatches: 0
XX Query Match: 99.6% Indels: 0
XX DB: 2 Gaps: 0
XX
XX US-10-718-311-4 (1-165) x AAT29820 (1-2000)
XX
QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db ATGTCACACCGCGTTAACGCAACTGCGTGGCTATGTAAGAGATCCCTGCC 439
QY 21 LeuAspProGlnLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGA 499
QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db CAGCAGGGGAAACCGGTAAAGCTGACGATGATCGCGAAGGGTTGTGCGAGCAGATGAA 559
QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db ATCCCGGAGAACTGCGCGTCTGCGGAAAGAGTCTCGTTACTGGTTACGTCAAAATTTG 619
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db TTATGTGCGGATGGTGAACCGTGGCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 679
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db ACCGGCGGAGCTGGCTTACAAAATTTGGGTAAACGCCGTTAGGACGCTATCTGTTTC 739
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db ACATCATCGACATTAACCGGACCTTTATTGAGATAGGCGGTGATCGCGGCTGTGGGG 799
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db CGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTACCG 859

```

```

QY 161 AlaSerProLeuTyr 165
Db 860 GCGTCACCGTTGTAC 874
XX
XX RESULT 14
XX ADV91643
XX ID ADV91643 standard; DNA; 1971 BP.
XX AC ADV91643;
XX
XX 10-MAR-2005 (first entry)
XX
XX C. parapsilosis pHB 1-H- E. coli CPL chimeric DNA.
XX
XX Hydroquinone glucoside; arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside;
XX chorismate pyruvate lyase; CPL; 4-hydroxycinnamoyl-CoA hydratase/lyase;
XX HCHL; gene expression; hydroquinone; antiseptic; transgenic plant;
XX genetically engineered microorganism; antioxidant; antimicrobial agent;
XX anti-inflammatory agent; carcinogenic inhibitor; melanoma; cytostatic;
XX neoplasm; pHBA 1-hydroxylase; pHBA; para-hydroxybenzoic acid;
XX p-hydroxybenzoic acid; gene; ds.
XX
XX Candida parapsilosis.
XX OS Escherichia coli.
XX Chimeric.
XX
XX Key Location/Qualifiers
XX CDS 4..501
XX /*tag= a
XX /product= "Escherichia coli CPL protein"
XX /transl_except= (pos: 46..48, aa:Cys)
XX 522..1561
XX /*tag= b
XX /product= "Candida parapsilosis pHBA 1-H protein"
XX /transl_except= (pos: 540..542, aa:Trp)
XX
XX US2004261147-A1.
XX
XX 23-DEC-2004.
XX
XX 16-JUN-2003; 2003US-00462162.
XX
XX 16-JUN-2003; 2003US-00462162.
XX
XX (MEYE/) MEYER K.
XX (VIIT/) VIITANEN P V.
XX (FLIN/) FLINT D.
XX
XX Meyer K, Viitanen PV, Flint D;
XX
XX WPI; 2005-057232/06.
XX P-PSDB; ADV91605, ADV91619.
XX
XX Producing hydroquinone glucoside in a green plant comprises growing a
XX green plant having nucleic acid fragments and chorismate pyruvate lyase
XX expression cassette.
XX
XX Example 7; SEQ ID NO 54; 70pp; English.
XX
XX The present invention relates to methods and materials to produce
XX hydroquinone glucoside (arbutin; 4-hydroxyphenyl-beta-D-glucopyranoside)
XX in genetically modified green plants and microorganisms. The method
XX relies upon transgenic plants or genetically modified microorganisms that
XX produce increased levels of the initial substrate para-hydroxybenzoic
XX acid (p-hydroxybenzoic acid; pHBA) in a biosynthetic pathway to produce
XX arbutin. Plants and microbes can be genetically engineered to produce
XX high levels of pHBA either by functional expression of the bacterial
XX protein chorismate pyruvate lyase (CPL) or by the expression of 4-
XX hydroxycinnamoyl-CoA hydratase/lyase (HCHL) or a combination of both.
XX Arbutin is useful as an antioxidant, antimicrobial agent, anti-
XX inflammatory agent and inhibitor of carcinogenesis (melanoma). The
XX present sequence is Candida parapsilosis pHBA 1-hydroxylase (pHB 1-H) and

```



```
CC Escherichia coli chorismate pyruvate lyase (CPL) chimeric DNA.
XX Sequence 1971 BP; 550 A; 383 C; 484 G; 554 T; 0 U; 0 Other;
SQ Alignment Scores:
Pred. No.: 8,23e-87 Length: 1971
Score: 843.00 Matches: 164
Percent Similarity: 99.4% Conservatives: 0
Best Local Similarity: 99.4% Mismatches: 1
Query Match: 98.7% Indels: 0
DB: 14 Gaps: 0
US-10-718-311-4 (1-165) x ADV91643 (1-1971)
Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 4 ATGTCACACCCCGCGTTAAACCACTGCGTGGCTGCTATTATTAAGAGAGATCCCTGCC 63
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 64 CTGGATCCGCAACTGCTCGACTGGCTGTGTCTGGAGGATTCATGACAAAAACGTTTGA 123
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGluAsnGlu 60
Db 124 CAGCAGGGAANAACGGTAAGCGTAGCATGATCCCGAAGGGTTTGTCCGACGAGATGAA 183
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 184 ATCCCGCAAGAACTGCCGCTGCTGCCGAAGAGTCTCGTTACTGTTACGTGAAATTTG 243
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 244 TTATGTGCCGATGGTGAACCGTGGCTGCCCGTGTACCGTGTCTCCGTGTCAACGTTA 303
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 304 AGCGGCGGAGCTGGCGTTACAAAATTGGTAANAACGCCGTTAGGACGCTATCTGTT 363
Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 364 ACATCATCGACATTAAACCCGGGACTTTATTGAGATAGGCGGTGATGCCGGCTGTGGGG 423
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 424 CGAGCTTCCCGCTCGGATTAAGCGGTAACCGCTGTGCTTAACAGAACTGTTTTTACCG 483
Qy 161 AlaSerProLeuTyr 165
Db 484 GCGTCACCGTTGTAC 498
RESULT 15
AAS90335
ID AAS90335 standard; cDNA; 1207 BP.
XX AAS90335;
AC AAS90335;
XX 13-FEB-2002 (first entry)
DT DNA encoding novel human diagnostic protein #26139.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
PN 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
PF 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
PR
```

---

```
XX (HYSE-) HYSEQ INC.
FA Drmanac RT, Liu C, Tang YT;
PI WPI: 2001-639362/73.
XX P-PSDB; ABG26148.
DR New isolated polynucleotide and encoded polypeptides, useful in
DR diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 1; SEQ ID NO 26139; 103pp; English.
PS The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS84197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 1207 BP; 316 A; 265 C; 288 G; 338 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2.06e-83 Length: 1207
Score: 811.00 Matches: 163
Percent Similarity: 98.2% Conservatives: 1
Best Local Similarity: 97.8% Mismatches: 1
Query Match: 95.0% Indels: 2
DB: 5 Gaps: 0
US-10-718-311-4 (1-165) x AAS90335 (1-1207)
Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 652 ATGTCACACCCCGCGTTAAACCACTGCGTGGCTGCTATTATTAAGAGATCCCTGCC 711
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 712 CTGGATCCGCAACTGCTCGACTGGCTGTGTCTGGAGGATTCATGACAAAAACGTTTGA 771
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGluAsnGlu 60
Db 772 CAGCAGGGAANAACGGTAAGCGTAGCATGATCCCGAAGGGTTTGTCCGACGAGATGAA 831
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 832 ATCCCGCAAGAACTGCGCTGCTCCGAAAGAGTCTCGTTACTGCTTACGTGAAATTTTG 891
Qy 81 -LeuCysAlaAspGlyGluProTrpLeuAla-GlyArgThrValProValSerThrL 100
Db 892 TGTATGTCCGATGGCGAACC CGCGCTTGC CGGTCGTACCGTCTGTTCTGTGTCAACGT 951
Qy 100 euSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuP 120
Db 952 TAACGGGCGGAGCTGCGGTTACAAAAATGGGGTAAAAACCCCGTAGGACGCTATCTGT 1011
```



QY 120 heThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpG 140  
Db |||||||  
1012 TCACATCATCGACATTAAACCGGACTTTATTGAGATAGCCCGTGATGCCGGCTGTGGG 1071  
QY 140 lyArgArgSerArgLeuArgLeuSerGlyIysProLeuLeuThrGluLeuPheLeup 160  
Db |||||||  
1072 GGCGACGTTCCGCGCTTGCATTAAAGCGGTAAACCGCTGTGTCTAACAGAACTGTTTTTAC 1131  
QY 160 roAlaSerProLeuTyr 165  
Db |||||||  
1132 CGCGGTCACCGTTGTAC 1148

Search completed: May 30, 2006, 02:42:29  
Job time : 730.015 secs

**THIS PAGE BLANK (uspto)**

GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 30, 2006, 01:30:02 ; Search time 5000.73 Seconds  
(without alignments)

3164.934 Million cell updates/sec

Title: US-10-718-311-4

Perfect score: 854

Sequence: 1 MSHPALQALRYCKEIPA.....RLSGRLLLTLELPASPLY 165

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 6366136 seqs, 31973710525 residues

Total number of hits satisfying chosen parameters: 12732272

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DRV=xlh  
-Q=abes/ABSWEB.spool/US10718311/runat\_26052006\_164916\_13439/app.query.fasta\_1  
-DB=GenEmbl -OFT=fastap -SUFFIX=rg -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abes03h  
-USER=US10718311 @CGN\_1\_1\_5767@runat\_26052006\_164916\_13439 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*

- 1: gb\_env.\*
- 2: gb\_pat.\*
- 3: gb\_ph.\*
- 4: gb\_pl.\*
- 5: gb\_pr.\*
- 6: gb\_ro.\*
- 7: gb\_sts.\*
- 8: gb\_sy.\*
- 9: gb\_un.\*
- 10: gb\_vi.\*
- 11: gb\_ov.\*
- 12: gb\_htg.\*
- 13: gb\_in.\*
- 14: gb\_om.\*
- 15: gb\_ba.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	854	100.0	495	2	A47932 Sequence 1
2	854	100.0	495	2	AR428733 Sequence
3	854	100.0	495	2	AR455409 Sequence

4	854	100.0	495	2	AX329364	AX329364 Sequence
5	854	100.0	495	2	AX382254	AX382254 Sequence
6	854	100.0	684	2	AR428736	AR428736 Sequence
7	854	100.0	684	2	AR455412	AR455412 Sequence
8	854	100.0	684	2	AX329368	AX329368 Sequence
9	854	100.0	684	2	AX382258	AX382258 Sequence
10	854	100.0	854	2	AR144788	AR144788 Sequence
11	854	100.0	1498	15	ECUBIA	E11274 ubiC and ub
12	854	100.0	2000	2	E11274	M93136 Escherichia
13	854	100.0	2000	15	ECUBICA	M93413 Escherichia
14	854	100.0	2000	15	ECUBIPLS	X57434 E.coli ubiC
15	854	100.0	2034	15	ECUBIAC	X66619 E.coli gene
16	854	100.0	2348	15	ECUBI	AR144789 Sequence
17	854	100.0	6641	2	AR144789	Continuation (43 o
18	854	100.0	110000	15	U00096_42	Continuation (43 o
19	854	100.0	110000	15	AP009048_42	U00006 E. coli chr
20	854	100.0	176195	15	ECUM89_42	Continuation (52 o
21	850	99.5	110000	15	AE005174_51	Continuation (51 o
22	846	99.1	110000	15	BA000007_50	Continuation (52 o
23	846	99.1	110000	15	BA000007_51	Continuation (52 o
24	843	98.7	1396	15	D0087228_42	D0087228 Escherich
25	843	98.7	110000	15	AE005674_43	Continuation (44 o
26	843	98.7	110000	15	AE014073_34	Continuation (35 o
27	843	98.7	110000	15	CP000036_41	Continuation (42 o
28	839	98.2	110000	15	CP000034_42	Continuation (43 o
29	835	97.8	110000	15	CP000038_44	Continuation (45 o
30	833	97.5	110000	15	AE014075_47	Continuation (48 o
31	798	93.4	1918	2	E07849	E07849 gDNA encodi
32	725	84.9	23880	15	AE008898	AE008898 Salmonell
33	725	84.9	110000	15	CP000026_41	Continuation (42 o
34	725	84.9	110000	15	CP000026_42	Continuation (43 o
35	721	84.4	110000	15	AE014613_42	Continuation (43 o
36	721	84.4	110000	15	AE017220_43	Continuation (44 o
37	721	84.4	234050	15	AL627282	AL627282 Salmonell
38	673	78.8	645	2	AR384505	AR384505 Sequence
39	437.5	51.2	10668	15	AE013658	AE013658 Yersinia
40	437.5	51.2	110000	15	BX936398_04	Continuation (5 of
41	437.5	51.2	214050	15	AJ414142	AJ414142 Yersinia
42	437.5	51.2	290803	15	AE017128	AE017128 Yersinia
43	424.5	49.7	110000	15	BX950851_06	Continuation (7 of
44	417	48.8	349107	15	BX571873	BX571873 Photorhab
45	408.5	47.8	110000	15	BX897699_12	Continuation (13 o

ALIGNMENTS

RESULT 1	A47932	Sequence 1 from Patent WO9600788.	495 bp	DNA	linear	PAT 07-MAR-1997
LOCUS	A47932	Sequence 1 from Patent WO9600788.				
DEFINITION	A47932					
ACCESSION	A47932					
VERSION	A47932.1	GI:2301795				
KEYWORDS						
SOURCE						
ORGANISM						
	Escherichia coli					
	Escherichia coli					
	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;					
	Enterobacteriaceae; Escherichia.					
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
COMMENT						
	Heide, L., Severin, K. and Siebert, M.					
	TRANSGENIC PLANTS WITH AN INCREASED SECONDARY SUBSTANCE CONTENT					
	Patent: WO 9600788-A 1 11-JAN-1996;					
	HEIDE LUTZ (DE)					
	Other publication AU 2879295 960125					
	Other publication DE 4423022 950524.					
FEATURES						
	Location/Qualifiers					
	1..495					
	/organism="Escherichia coli"					
	/mol_type="unassigned DNA"					
	/strain="MC4100"					
	/db_xref="taxon:562"					

source

ORIGIN

Alignment Scores:

```
Pred. No.: 1.54e-81 Length: 495
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 2 Gaps: 0

US-10-718-311-4 (1-165) x A47932 (1-495)
Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluileProAla 20
Db 1 ATGTCACACCCCGCGTTAACGCAACTGCGTGGCTGCGCTATTGTAAAGAGATCCCTGCC 60
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATCCATGACAAAAACGTTTGA 120
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 121 CAGCAGGAGAAAAACGGTAAGCGTACCATGATCCGCAAGGGTTTGTGACAGAGATGAA 180
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluileLeu 80
Db 181 ATCCCCGAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTG 240
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 241 TTATGTCCGATGGTGAACCGTGGCTTACAAAAATTGGTAAACGCCGTAGGACGCTAICTGTT 360
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 301 AGCGGCGCGAGCTGGCGTTACAAAAATTGGTAAACGCCGTAGGACGCTAICTGTT 360
Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluileGlyArgAspAlaGlyLeuTrpGly 140
Db 361 ACATCATCGACATTAAACCCGGGACTTTATTGAGATAGCGCGTATGCGGGCTGTGGGG 420
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 160
Db 421 CGAGCTTCCCGCTCGGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTACCG 480
Qy 161 AlaSerProLeuTyr 165
Db 481 GCGTCACCGTTGTAC 495

RESULT 3
AR455409 AR455409 495 bp DNA linear PAT 20-FEB-2004
LOCUS Sequence 3 from patent US 6683231.
DEFINITION AR455409
ACCESSION AR455409
VERSION AR455409.1 GI:42689957
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 495)
AUTHORS Meyer,K., Viitanen,P.V. and Van Dyk,D.E.
TITLE High level production of p-hydroxybenzoic acid in green plants
JOURNAL Patent: US 6683231-A 3 27-JAN-2004;
E. I. du Pont de Nemours and Company; Wilmington, DE
FEATURES
source Location/Qualifiers
1..495
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores: 1.54e-81 Length: 495
Pred. No.: 854.00 Matches: 165
Score: 100.0% Conservative: 0
Percent Similarity: 100.0% Mismatches: 0
Best Local Similarity: 100.0% Indels: 0
Query Match: 100.0% Gaps: 0
DB: 2

US-10-718-311-4 (1-165) x AR455409 (1-495)
Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluileProAla 20
Db 1 ATGTCACACCCCGCGTTAACGCAACTGCGTGGCTGCGCTATTGTAAAGAGATCCCTGCC 60
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATCCATGACAAAAACGTTTGA 120
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 121 CAGCAGGAGAAAAACGGTAAGCGTACCATGATCCGCAAGGGTTTGTGACAGAGATGAA 180
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluileLeu 80
Db 181 ATCCCCGAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTG 240
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 241 TTATGTCCGATGGTGAACCGTGGCTTACAAAAATTGGTAAACGCCGTAGGACGCTAICTGTT 360
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 301 AGCGGCGCGAGCTGGCGTTACAAAAATTGGTAAACGCCGTAGGACGCTAICTGTT 360
Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluileGlyArgAspAlaGlyLeuTrpGly 140
Db 361 ACATCATCGACATTAAACCCGGGACTTTATTGAGATAGCGCGTATGCGGGCTGTGGGG 420
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 160
Db 421 CGAGCTTCCCGCTCGGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTACCG 480
Qy 161 AlaSerProLeuTyr 165
Db 481 GCGTCACCGTTGTAC 495

RESULT 2
AR428733 AR428733 495 bp DNA linear PAT 18-DEC-2003
LOCUS Sequence 12 from patent US 6642036.
DEFINITION AR428733
ACCESSION AR428733
VERSION AR428733.1 GI:40188463
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 495)
AUTHORS Flint,D., Meyer,K. and Viitanen,P.
TITLE Sinapoyglucose:malate sinapoyltransferase form malate conjugates
JOURNAL Patent: US 6642036-A 12 04-NOV-2003;
E. I. du Pont de Nemours and Company; Wilmington, DE
FEATURES
source Location/Qualifiers
1..495
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores: 1.54e-81 Length: 495
Pred. No.: 854.00 Matches: 165
Score: 100.0% Conservative: 0
Percent Similarity: 100.0% Mismatches: 0
Best Local Similarity: 100.0% Indels: 0
Query Match: 100.0% Gaps: 0
DB: 2

US-10-718-311-4 (1-165) x AR428733 (1-495)
Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluileProAla 20
Db 1 ATGTCACACCCCGCGTTAACGCAACTGCGTGGCTGCGCTATTGTAAAGAGATCCCTGCC 60
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATCCATGACAAAAACGTTTGA 120
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 121 CAGCAGGAGAAAAACGGTAAGCGTACCATGATCCGCAAGGGTTTGTGACAGAGATGAA 180
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluileLeu 80
Db 181 ATCCCCGAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTG 240
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 241 TTATGTCCGATGGTGAACCGTGGCTTACAAAAATTGGTAAACGCCGTAGGACGCTAICTGTT 360
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 301 AGCGGCGCGAGCTGGCGTTACAAAAATTGGTAAACGCCGTAGGACGCTAICTGTT 360
Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluileGlyArgAspAlaGlyLeuTrpGly 140
Db 361 ACATCATCGACATTAAACCCGGGACTTTATTGAGATAGCGCGTATGCGGGCTGTGGGG 420
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 160
Db 421 CGAGCTTCCCGCTCGGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTACCG 480
Qy 161 AlaSerProLeuTyr 165
Db 481 GCGTCACCGTTGTAC 495
```

```
Db      1  ATGTCTACACCCCGCTTAACGCAACTGCGTGCCTGCGCTATTTGTAAGAGATCCCTGCC 60
Qy      21  LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db      61  CTGGATCCCGAACTGCTCAGCTGCTGTTGCTGGAGATTCATGACAAACGTTTGA 120
Qy      41  GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAenGlu 60
Db      121  CAGCAGGGGAAAACCGTAGCGTACGATGATCCGGAAGGGTTTGTTCGAGCAGATGAA 180
Qy      61  IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeuLeu 80
Db      181  ATCCCGCGAAGAACTGCGCGTGTGCGGAAAGAGTCTCGTTACTGTTAGTGAAAATTTG 240
Qy      81  LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db      241  TTATGTGCCGATGGTGAACCGTGGCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
Qy      101  SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db      301  AGCGGCGCGAGCTGCGCTTACAAAATTTGGTAAACCGCTTAGGACGCTATCTGTTTC 360
Qy      121  ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db      361  ACATCATCGACATTAAACCGGACTTTATTGAGATAGGCGGTGATGCCGGCTGTGGGG 420
Qy      141  ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db      421  CGACGTTCCCGCTCGGATTAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTACCG 480
Qy      161  AlaSerProLeuTyr 165
Db      481  GCGTCACCGTTGTAC 495
```

```
RESULT 4
AX329364
LOCUS      AX329364
DEFINITION Sequence 3 from Patent WO0194607.
ACCESSION AX329364
VERSION    AX329364.1 GI:18102364
KEYWORDS   .
SOURCE     unidentified
ORGANISM   unidentified
REFERENCE  1
AUTHORS    Meyer, K., van Dyk, D.E. and Viitanen, P.V.
TITLE      High level production of p-hydroxybenzoic acid in green plants
JOURNAL    Patent: WO 0194607-A 3 13-DEC-2001;
           E.I. Dupont De Nemours (US)
FEATURES   Location/Qualifiers
            source
              1..495
              /organism="unidentified"
              /mol_type="unassigned DNA"
              /db_xref="taxon:32644"
              /note="Description of Unknown Organism:E. coli"
```

```
ORIGIN
Alignment Scores:
Pred. No.:      1.54e-81      Length:      495
Score:          854.00      Matches:      165
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    100.0%      Indels:      0
DB:             2           Gaps:          0
US-10-718-311-4 (1-165) x AX329364 (1-495)
```

```
Qy      1  MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db      1  ATGTACACCCCGCTTAACGCAACTGCGTGCCTATTTGTAAGAGATCCCTGCC 60
Qy      21  LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
```

```
Db      61  CTGGATCCCGCAACTGCTCAGCTGGCTGTTGCTGGAGATTCCATGACAAAACGTTTGA 120
Qy      41  GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAenGlu 60
Db      121  CAGCAGGGGAAAACCGTAGCGTACGATGATCCGGAAGGGTTTGTTCGAGCAGATGAA 180
Qy      61  IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db      181  ATCCCGCGAAGAACTGCGCGTGTGCGGAAAGAGTCTCGTTACTGTTAGTGAAAATTTG 240
Qy      81  LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db      241  TTATGTGCCGATGGTGAACCGTGGCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
Qy      101  SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db      301  AGCGGCGCGAGCTGCGCTTACAAAATTTGGTAAACCGCTTAGGACGCTATCTGTTTC 360
Qy      121  ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db      361  ACATCATCGACATTAAACCGGACTTTATTGAGATAGGCGGTGATGCCGGCTGTGGGG 420
Qy      141  ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db      421  CGACGTTCCCGCTCGGATTAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTACCG 480
Qy      161  AlaSerProLeuTyr 165
Db      481  GCGTCACCGTTGTAC 495
```

```
RESULT 5
AX382254
LOCUS      AX382254
DEFINITION Sequence 12 from Patent WO0204653.
ACCESSION AX382254
VERSION    AX382254.1 GI:19577034
KEYWORDS   .
SOURCE     Escherichia coli
ORGANISM   Escherichia coli
REFERENCE  1
AUTHORS    Flint, D., Meyer, K. and Viitanen, P.V.
TITLE      Sinapolyglucose:malate sinapolytransferase form malate conjugates
           from benzoic acid glucosides
JOURNAL    Patent: WO 0204653-A 12 17-JAN-2002;
           E.I. DUPONT DE NEMOURS AND COMPANY (US)
FEATURES   Location/Qualifiers
            source
              1..495
              /organism="Escherichia coli"
              /mol_type="unassigned DNA"
              /db_xref="taxon:562"
```

```
ORIGIN
Alignment Scores:
Pred. No.:      1.54e-81      Length:      495
Score:          854.00      Matches:      165
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    100.0%      Indels:      0
DB:             2           Gaps:          0
US-10-718-311-4 (1-165) x AX382254 (1-495)
```

```
Qy      1  MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db      1  ATGTACACCCCGCTTAACGCAACTGCGTGCCTATTTGTAAGAGATCCCTGCC 60
Qy      21  LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db      61  CTGGATCCCGCAACTGCTCAGCTGGCTGTTGCTGGAGGATTCCATGACAAAACGTTTGA 120
```

```
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 121 CAGCAGGGAAGAAACGGTAAGCGTGACGATGATCCGGAAGGGTTGTCTGACGACAGAAATGAA 180

Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 181 ATCCCGGAAGAACTGCGCTGCTGCCGAAGAGTCTCGTTACTGGTACGTGAATTTTG 240

Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db 241 TTATGTGCGGATGGTGAACCGTGGCTTGCCGTCGTACCGTCTCTCTGTGTCAACGTTA 300

Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 301 AGCGGCGCGAGCTGGCGCTTACAAAATTGGTAAACCGCGGTAGGACGCTATCTGTTC 360

Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 361 ACATCATCGACATTAAACCGGACTTATTAGATAGCCCGTATGCCGGCTGTGGGG 420

Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 421 CGACGTTCCCGCTCGCATTAAGCGTTAAACCGCTGTGCTAACAGAACTGTTTTACCG 480

Qy 161 AlaSerProLeuTyr 165
Db 481 GCGTCACCGTTGTAC 495

RESULT 6
AR428736
LOCUS AR428736 684 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 16 from patent US 6642036.
ACCESSION AR428736
VERSION AR428736.1 GI:40188466
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 684)
AUTHORS Flint,D., Meyer,K. and Viitanen,P.
TITLE Sinapoylglucose:malate sinapoyltransferase form malate conjugates
JOURNAL Patent: US 6642036-A 16 04-NOV-2003;
E. I. Du Pont de Nemours and Company; Wilmington;
WOX;

FEATURES
source Location/Qualifiers
1..684
/mol_type="genomic DNA"

ORIGIN
Alignment Scores: 2.19e-81 Length: 684
Pred. No.: 854.00 Matches: 165
Score: 100.0% Conservative: 0
Percent Similarity: 100.0% Mismatches: 0
Best Local Similarity: 100.0% Indels: 0
Query Match: 100.0% Gaps: 0
DB: 2

US-10-718-311-4 (1-165) x AR428736 (1-684)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 187 ATGTCACACCCCGGCTTAACGCAACTGCGTGGCTGCTATTGTAAGAGATCCCTGCC 246

Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 247 CTGGATCCGCAACTGCTCGA CTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGAA 306

Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 307 CAGCAGGGAAGAAACGGTAAGCGTGACGATATCCGGAAGGGTTGTTCGACAGAAATGAA 366

Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 367 ATCCCGGAAGAAACGGTAAGCGTGACGATATCCGGAAGGGTTGTTCGACAGAAATTTG 426

Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
```

```
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 367 ATCCCGGAAGAAACGGTGCCTGCCGAAGAGTCTCGTTACTGTTACGTGAATTTTG 426

Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db 427 TTATGTGCGGATGGTGAACCGTGGCTTGCCGTCGTACCGTCTCTCTGTGTCAACGTTA 486

Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 487 AGCGGCGCGAGCTGGCGTTACAAAATTGGTAAACCGCGTTAGGACGCTATCTGTTC 546

Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 547 ACATCATCGACATTAAACCGGACTTATTAGATAGCCCGTATGCCGGCTGTGGGG 606

Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 607 CGACGTTCCCGCTCGCATTAAGCGTTAAACCGCTGTGCTAACAGAACTGTTTTACCG 666

Qy 161 AlaSerProLeuTyr 165
Db 667 GCGTCACCGTTGTAC 681

RESULT 7
AR455412
LOCUS AR455412 684 bp DNA linear PAT 20-FEB-2004
DEFINITION Sequence 7 from patent US 6683231.
ACCESSION AR455412
VERSION AR455412.1 GI:42689960
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 684)
AUTHORS Meyer,K., Viitanen,P.V. and Van Dyk,D.E.
TITLE High level production of p-hydroxybenzoic acid in green plants
JOURNAL Patent: US 6683231-A 7 27-JAN-2004;
E. I. du Pont de Nemours and Company; Wilmington, DE

FEATURES
source Location/Qualifiers
1..684
/mol_type="genomic DNA"

ORIGIN
Alignment Scores: 2.19e-81 Length: 684
Pred. No.: 854.00 Matches: 165
Score: 100.0% Conservative: 0
Percent Similarity: 100.0% Mismatches: 0
Best Local Similarity: 100.0% Indels: 0
Query Match: 100.0% Gaps: 0
DB: 2

US-10-718-311-4 (1-165) x AR455412 (1-684)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 187 ATGTCACACCCCGGCTTAACGCAACTGCGTGGCTGCTATTGTAAGAGATCCCTGCC 246

Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 247 CTGGATCCGCAACTGCTCGA CTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGAA 306

Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 307 CAGCAGGGAAGAAACGGTAAGCGTGACGATATCCGGAAGGGTTGTTCGACAGAAATGAA 366

Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 367 ATCCCGGAAGAAACGGTGCCTGCCGAAGAGTCTCGTTACTGTTACGTGAATTTTG 426

Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
```

Db 427 TTATGTGCCGATGGTGAAACCGTGGCTTGGCGGTTCGTAACCGTCTTCTGTGTCAACGTTA 486  
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120  
Db 487 AGCGGCGCGAGCTGGCTTACAAAATTGGGTAAACCGCTTAGACGCTATCTGTTC 546  
Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140  
Db 547 ACATCATCGACATTAAACCGGACTTTATTGAGTAGGCGGTGATGCCGGCTGTGGGG 606  
Qy 141 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140  
Db 547 ACATCATCGACATTAAACCGGACTTTATTGAGTAGGCGGTGATGCCGGCTGTGGGG 606  
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160  
Db 607 CGACGTTCCCGCTCGGATTAAACCGGTAAACCGCTGTGTCTAACAGAACTGTTTTACCG 666  
Qy 161 AlaSerProLeuTyr 165  
Db 667 GCGTCACCGTTGTAC 681  
RESULT 8  
AX329368  
LOCUS AX329368 684 bp DNA linear PAT 08-JAN-2002  
DEFINITION Sequence 7 from Patent WO0194607.  
ACCESSION AX329368  
VERSION AX329368.1 GI:18102367  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1  
AUTHORS Meyer, K., van Dyk, D.E. and Viitanen, P.V.  
TITLE High level production of p-hydroxybenzoic acid in green plants  
JOURNAL Patent: WO 0194607-A 7 13-DEC-2001;  
E.I. Dupont De Nemours (US)  
FEATURES  
Location/Qualifiers  
1..684  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="synthetic CPL"  
ORIGIN  
Alignment Scores:  
Pred. No.: 2.19e-81 Length: 684  
Score: 854.00 Matches: 165  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 2 Gaps: 0  
US-10-718-311-4 (1-165) x AX329368 (1-684)  
Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20  
Db 187 ATGTGCACACCGCGCTTAAACCACTGGCTGGCTATGTGTAAAGAGATCCCTGCC 246  
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40  
Db 247 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGATTCCATGACAAAAGCTTTTGA 306  
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAsnGlu 60  
Db 307 CAGCAGGGAAAACCGTAAAGCGTACGATGATCCCGAAGGGTGTTCGAGCAGAAATGA 366  
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80  
Db 367 ATCCCCGGAAGAACTGCCGCTGTCGCGAAAGAGTCTCGTTACTGTGTACGTGAAAATTTG 426  
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100  
Db 427 TTATGTGCCGATGGTGAACCGTGGCTTGCCGGCTCGTACCGTCTCTGTGTCAACGTTA 486  
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120

Db 487 AGCGGCGCGAGCTGGCTTACAAAATTGGGTAAACCGCTTAGACGCTATCTGTTC 546  
Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140  
Db 547 ACATCATCGACATTAAACCGGACTTTATTGAGTAGGCGGTGATGCCGGCTGTGGGG 606  
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160  
Db 607 CGACGTTCCCGCTCGGATTAAACCGGTAAACCGCTGTGTCTAACAGAACTGTTTTACCG 666  
Qy 161 AlaSerProLeuTyr 165  
Db 667 GCGTCACCGTTGTAC 681  
RESULT 9  
AX382258  
LOCUS AX382258 684 bp DNA linear PAT 18-MAR-2002  
DEFINITION Sequence 16 from Patent WO0204653.  
ACCESSION AX382258  
VERSION AX382258.1 GI:19577037  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1  
AUTHORS Flint, D., Meyer, K. and Viitanen, P.V.  
TITLE Sinapolyglucose:malate sinapolytransferase form malate conjugates  
JOURNAL Patent: WO 0204653-A 16 17-JAN-2002;  
E.I. DUPONT DE NEMOURS AND COMPANY (US)  
FEATURES  
Location/Qualifiers  
1..684  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="open reading frame of the chloroplast-targeted CPL  
fusion protein"  
ORIGIN  
Alignment Scores:  
Pred. No.: 2.19e-81 Length: 684  
Score: 854.00 Matches: 165  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 2 Gaps: 0  
US-10-718-311-4 (1-165) x AX382258 (1-684)  
Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20  
Db 187 ATGTGCACACCGCGCTTAAACCACTGGCTGGCTATGTGTAAAGAGATCCCTGCC 246  
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40  
Db 247 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGATTCCATGACAAAAGCTTTTGA 306  
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAsnGlu 60  
Db 307 CAGCAGGGAAAACCGTAAAGCGTACGATGATCCCGAAGGGTGTTCGAGCAGAAATGA 366  
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80  
Db 367 ATCCCCGGAAGAACTGCCGCTGTCGCGAAAGAGTCTCGTTACTGTGTACGTGAAAATTTG 426  
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100  
Db 427 TTATGTGCCGATGGTGAACCGTGGCTTGCCGGCTCGTACCGTCTCTGTGTCAACGTTA 486  
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120  
Db 487 AGCGGCGCGAGCTGGCGTTACAAAATTGGGTAAACCGCTTAGACGCTATCTGTTC 546

```

Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 547 ACATCATCGACATTAAACCGGACTTTATTGAGATAGCGCGTATGCCGGGCTGTGGGG 606
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 607 CGAGTTCGCCGCTCGATTAGCGGTAAACCGCTGTGCTAACAGAACTGTTTTTACCG 666
Qy 161 AlaSerProLeuTyr 165
Db 667 GCGTCACCGTTGTAC 681
RESULT 10
ARI44788
LOCUS
DEFINITION Sequence 24 from patent US 6210937.
ACCESSION ARI44788
VERSION ARI44788.1 GI:15106655
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 854)
AUTHORS Ward, T.E., Watkins, C.S., Bulmer, D.K., Johnson, B.F. and
Amaratunga, M.
TITLE Development of genetically engineered bacteria for production of
selected aromatic compounds
JOURNAL Patent: US 6210937-A 24 03-APR-2001;
FEATURES
Location/Qualifiers
1..854
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN
Alignment Scores:
Pred. No.: 2,79e-81 Length: 854
Scores: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
Gaps: 0
DB: 2
US-10-718-311-4 (1-165) x ARI44788 (1-854)
Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 20
Db 321 ATGTCACACCCCGCTTAACGCAACTGCGTGCCTGCTATTGTAAAGAGATCCCTGCC 380
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 381 CTGGATCCGCAACTGCTCGACTGGCTGTGTGGAGGATTCATGACAAAACGTTTGA 440
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 441 CAGCAGGGGAAAACCGTAGCGTGACGATGATCCCGAAGGGTTGTGCGACAGATGAA 500
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeuLeu 80
Db 501 ATCCCGAAGAACTGCCGCTGCTGCCGAAGAGTCTCGTTACTGCTTACGTGAAATTTG 560
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 561 TTATGTGCGGATGGTGAACCGGTGGCTTCCCGTGTACCGTGTCTGTGTCAACGTTA 620
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 621 AGCGGGCCGAGCTGGCGTTACAAAAATGGTAAACGCCGTTAGGAGCGTATCTGTTT 680
Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 681 ACATCATCGACATTAAACCGGACTTTATTGAGATAGCGCGTATGCCGGGCTGTGGGG 740
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160

```

```

Db 741 CGAGTTCGCCGCTGCGATTAAACCGGTAAACCGCTGTGCTAACAGAACTGTTTTTACCG 800
Qy 161 AlaSerProLeuTyr 165
Db 801 GCGTCACCGTTGTAC 815
RESULT 11
ECOUBIA
LOCUS
DEFINITION Escherichia coli 4-hydroxybenzoate octaprenyl transferase (ubiA)
gene, 5' end and chorismate lyase (ubiC) gene, complete cds.
ACCESSION M96268.1 GI:347886
VERSION
KEYWORDS 4-hydroxybenzoate octaprenyl transferase; aerobic respiratory
deficiency; chorismate lyase; membrane-bound protein;
prenyltransferase; transferase; ubiA gene; ubiC gene; ubiquinone
biosynthesis.
SOURCE
ORGANISM Escherichia coli
Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 (bases 1 to 1498)
AUTHORS Wu, G., Williams, H.D., Gibson, P. and Poole, R.K.
TITLE Mutants of Escherichia coli affected in respiration: the cloning
and nucleotide sequence of ubiA, encoding the membrane-bound
p-hydroxybenzoate: octaprenyltransferase
JOURNAL J. Gen. Microbiol. 139 (Pt 8), 1795-1805 (1993)
PUBMED 8409922
FEATURES
Location/Qualifiers
1..1498
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/strain="K-12"
/db_xref="taxon:562"
/map="91.8 min"
/tissue_lib="cosmid"
<1..524
/notes="ORF; the sequence before nucleotide number 110
might not be genuine. Three Clai sites, which is the
enzyme used for cloning, were identified in this region,
so it could be a cloning artifact"
/codon_start=3
/transl_table=11
/protein_id="AA17026.1"
/db_xref="GI:347887"
/transl_table=11
/translation="RFSEVNFDTQOMIDFINRFTNKWLIDKNNFFIQLIDQALRST
DDMIKANVHLKWEIRSDVSPFIETEDNLTNTFNLTNRDNIILFSSVDDGPV
MVVSSORLHDLNAPTCDTNNSTVIYKSRHEMLPVNLTQETLPSKSHGKVALPFIPT
ASWRARIMNKGV"
RBS
735..738
gene
747..1244
/gene="ubiC"
747..1244
CDS
/gene="ubiC"
/codon_start=1
/transl_table=11
/product="chorismate lyase"
/protein_id="AA17027.1"
/db_xref="GI:347888"
/translation="MSHPALTOLRALRYCKEIPALDPQLLDWLLLEDSTMTRFEOQKG
TVSVTMIREGFVQNEIPELPLLPKESRYLREILLCADGEPLHAGRTVVPVSTLGS
PEALQKLGKTLPLGYLFTSTLTRDFEIGRDAGLWGRSRLSLSGKPLLLTBLFLP
ASPLY"
1257..1498
gene
/gene="ubiA"
1257..1498
CDS
/gene="ubiA"
/notes="The last part of this ORF is to be found in entry
K00127.; ORF"
/codon_start=1
/transl_table=11
/product="4-hydroxybenzoate-octaprenyl transferase"

```



/protein\_id="AA17028.1"  
/db\_xref="GI:347889"  
/translation="MWSLTQNKLLAFHRLMRDTPKIGALLLLWPTLMALWATPGVP  
QLWILAVFVAGVWLMRAAGCVNDYADRFEDGHVKT"

## ORIGIN

Alignment Scores:  
Pred. No.: 5,13e-81 Length: 1498  
Score: 854.00 Matches: 165  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 15 Gaps: 0

US-10-718-311-4 (1-165) x ECOUBIA (1-1498)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20  
DB 747 ATGTACACCCCGCTTAACGCAACTGCGTGGCTGCTATTGTAAGAGATCCCTGCC 806  
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40  
DB 807 CTGGATCCGCAACTGCTGCACTGGCTGCTGGAGATCCATGACAAACGTTTGA 866  
QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAenGlu 60  
DB 867 CAGCAGGGAACGTAAGCGTGACGATGATCCGCAAGGGTTTCTCGACGAGATGAA 926  
QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80  
DB 927 ATCCCGCAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGTTAGCGTAAATTTG 986  
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100  
DB 987 TTATGTCCGATGGTGAACCGTGGCTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1046  
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120  
DB 1047 ACSCGGCGGAGCTGGCGTTACAAAATTGGGTAAACCGCTAGGACGCTATCTGTTTC 1106  
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140  
DB 1107 ACATCATCGACATTACCCGGGACTTTATTGATAGGCGGTGATGCGCGGCTGTGGGG 1166  
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrClnLeuPheLeuPro 160  
DB 1167 CGACGTTCCCGCTGGGATTAAGCGGTAAACCGCTGCTGCTAACAGAACTGTTTTACCG 1226  
QY 161 AlaSerProLeuTyr 165  
DB 1227 GCGTCACCGTTGTAC 1241

## RESULT 12

E11274  
LOCUS ubiC and ubiA gene.  
DEFINITION E11274  
ACCESSION E11274  
VERSION E11274.1 GI:22024916  
KEYWORDS JP 1996107789-A/1.  
SOURCE Escherichia coli  
ORGANISM Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
1 (bases 1 to 2000)  
REFERENCE Matsuda,H., Kawamuki,M. and Nakagawa,T.  
AUTHORS PRODUCTION OF UBIQUINONE-10  
TITLE Patent: JP 1996107789-A 1 30-APR-1996;  
JOURNAL ALPHA- SHOKUHIN KK  
COMMENT OS Escherichia coli  
PN JP 1996107789-A/1  
PD 30-APR-1996  
PF 13-OCT-1994 JP 1994273071  
PI MATSUDA HIDEYUKI, KAWAMUKI MAKOTO, NAKAGAWA TSUYOSHI PC

C12N15/09 A61K31/12 A61K31/12 C07H21/04 C12N1/21 C12P7/66 PC  
(C12N1/21,  
PC C12R1.01), (C12P7/66, C12R1.01);  
CC strandedness: Double;  
CC topology: Linear;  
FH Key Location/Qualifiers  
FT source 1..2000  
FT /organism='Escherichia coli'  
FT CDS 380..877  
FT /product='UbiC protein'  
FT CDS 890..1762  
FT /product='UbiA protein'.  
FEATURES  
Location/Qualifiers  
1..2000  
/organism='Escherichia coli'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:562'

## ORIGIN

Alignment Scores:  
Pred. No.: 7.02e-81 Length: 2000  
Score: 854.00 Matches: 165  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 2 Gaps: 0

US-10-718-311-4 (1-165) x E11274 (1-2000)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20  
DB 380 ATGTACACCCCGCTTAACGCAACTGCGTGGCTGCTATTGTAAGAGATCCCTGCC 439  
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40  
DB 440 CTGGATCCGCAACTGCTGCACTGGCTGCTGGAGATCCATGACAAACGTTTGA 499  
QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAenGlu 60  
DB 500 CAGCAGGGAACGTAAGCGTGACGATGATCCGCAAGGGTTTCTCGACGAGATGAA 559  
QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80  
DB 560 ATCCCGCAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGTTAGCGTAAATTTG 619  
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100  
DB 620 TTATGTCCGATGGTGAACCGTGGCTTGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 679  
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120  
DB 680 ACSCGGCGGAGCTGGCGTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 739  
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140  
DB 740 ACATCATCGACATTACCCGGGACTTTATTGATAGGCGGTGATGCGCGGCTGTGGGG 799  
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrClnLeuPheLeuPro 160  
DB 800 CGACGTTCCCGCTGGGATTAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTACCG 859  
QY 161 AlaSerProLeuTyr 165  
DB 860 GCGTCACCGTTGTAC 874

## RESULT 13

ECOUBICA  
LOCUS Escherichia coli chorismate lyase (ubiC), 4-hydroxybenzoate  
DEFINITION octaprenyl transferase (ubiA) genes, complete cds, and  
sn-glycerol-3-phosphate acyltransferase (plab) genes, 3' end.  
ACCESSION M93136

```

VERSION M93136.1 GI:148099
KEYWORDS 4-hydroxybenzoate octaprenyl transferase; chorismate lyase; plbB
          gene; sn-glycerol-3-phosphate acyltransferase; ubiA gene; ubiC
          gene.
SOURCE   Escherichia coli W3110
ORGANISM Escherichia coli W3110
          Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
          Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 2000)
AUTHORS Lightner, V.A., Bell, R.M. and Modrich, P.
TITLE The DNA sequences encoding plbB and dgk loci of Escherichia coli
JOURNAL J. Biol. Chem. 258 (18), 10856-10861 (1983)
PUBMED 6309817
REFERENCE 2 (bases 1 to 2000)
AUTHORS Nichols, B.P. and Green, J.M.
TITLE Cloning and Sequencing of Escherichia coli ubiC and purification of
          chorismate lyase
JOURNAL Unpublished (1992)
COMMENT Original source text: Escherichia coli (sub_strain W3110, strain
          K-12) (library: Kohare lambda-1F8) DNA.
          Location/Qualifiers
FEATURES             source
1..2000
/organism="Escherichia coli W3110"
/mol_type="genomic DNA"
/strain="K-12"
/sub_strain="W3110"
/db_xref="taxon:316407"
/map="91.8 min"
/tissue_lib="Kohare lambda-1F8"
380..877
/gene="ubiC"
380..877
/gene="ubiC"
/function="enzymatic chorismate --> p-hydroxybenzoate +
pyruvate"
/experiment="experimental evidence, no additional details
recorded"
/citation=[2]
/codon_start=1
/transl_table=11
/product="chorismate lyase"
/protein_id="AAA24711.1"
/db_xref="GI:148100"
/translation="MSHPALTQLRALRYCKEIPALDQLLDWLLLESMTKPRQOQK
TVSTMREGVEQNEIPEDPLPKESRYLREILLCADGEPMLAGRTVVPVSTLSG
PELALQKLKGTPLGRYUFTSTLTRDFEIGRDAGLWRRSLRSLKPLLLTEFLFP
ASPLY"
890..1762
/gene="ubiA"
890..1762
/gene="ubiA"
/function="enzymatic - 3-octaprenyl-4-hydroxybenzoate
synthesis"
/experiment="experimental evidence, no additional details
recorded"
/citation=[2]
/codon_start=1
/transl_table=11
/product="4-hydroxybenzoate-octaprenyl transferase"
/protein_id="AAA24712.1"
/db_xref="GI:148101"
/translation="MWSLTONKLLAFHRLMRTDKPIGALLLLNPTLWALWVATPGVP
QLWILAVFAGVGLMRAAGCVVNDYDRKFDGHRVKTANRPLPSGAVTEREARALFVV
LVLSFLDLVTNTMTLLSALALAWVPFKRYTHLPQVILGAAPFGWSIPMAFAA
VSEVSPCLMFLANILWAVYDTQVAMVDRDDDKIKSTAILRFQYDKLIJGIL
QIGVALMAIIIGELNGLGWYSILVAGALFVYQOKLIANREREAFCFAPMNNYVG
LVFLGLAWSYHF"
complement(1917..2000)
/gene="plbB"
complement(1917..2000)
/gene="plbB"
/EC_number="2.3.1.15"
/function="lysophosphatidic acid synthesis"

/experiment="experimental evidence, no additional details
recorded"
/citation=[1]
/codon_start=1
/transl_table=11
/product="sn-glycerol-3-phosphate acyltransferase"
/protein_id="AAA24713.1"
/db_xref="GI:148102"
/translation="MKVIQLLAEILITSDVRLTIESATQGEG"

ORIGIN
Alignment Scores:
Pred. No.: 7,02e-81 Length: 2000
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservat: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 15 Gaps: 0

US-10-718-311-4 (1-165) x ECOUBICA (1-2000)
Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgfyrCysLysGluileProAla 20
Db 380 ATGTACACCCCGCGTTAAACGCAACTGCTGCGCTGCTATTGTTAAAGAGATCCCTGCC 439
Qy 21 LeuaspProGlnLeuLeuaspTrpLeuLeuLeuGluaspSerMetThrLysArgPheGlu 40
Db 440 CTGGATCCGCAACTGCTGCACTGGCTGTGCTGAGGATTCATGACAAAAACGTTTGA 499
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db 500 CAGCAGGGAACCGTTAAGCGTACGATGATCCGCAAGGGTTTGTTCGACAGAAATGAA 559
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgfyrTrpLeuArgGluileLeu 80
Db 560 ATCCCGCAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGCTTACGTGAAATTTG 619
Qy 81 LeuCysAlaaspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db 620 TTATGTCCGATGGTGAACCGTGGCTTGGCGGTCGTACCGTCGTCTCTGTGTCAACGTTA 679
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgfyrLeuPhe 120
Db 680 AGCGGGCGGAGCTGGCGGTACAAAATTTGGTAAAACCGCGTTAGGACGCTATCTGTTTC 739
Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluileGlyArgAspAlaGlyLeuTrpGly 140
Db 740 ACATCATCGACATTAAACCCGGACTTTATTTCAGATAGCCCGCTGATGCCGGCTGTGGGG 799
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 800 CGAGTTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTGCTTACAGAACTGTTTTTACCG 859
Qy 161 AlaSerProLeuTyr 165
Db 860 GCGTCACCGTTGTAC 874

RESULT 14
ECOUBIPLS 2000 bp DNA linear BCT 26-APR-1993
LOCUS Escherichia coli 4-hydroxybenzoate octaprenyl transferase (ubiA)
DEFINITION gene complete cds, chorismate lyase (ubiC) gene complete cds,
          sn-glycerol-3-phosphate acyltransferase (plbB) gene, 3' end.
          M93413
ACCESSION M93413.1 GI:148106
VERSION 4-hydroxybenzoate octaprenyl transferase; chorismate lyase; plbB
          gene; sn-glycerol-3-phosphate acyltransferase; ubiA gene; ubiC
          gene.
SOURCE Escherichia coli W3110
ORGANISM Escherichia coli W3110
          Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
          Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 2000)

```

/product="en-glycerol-3-phosphate acyltransferase"  
 /protein\_id="AAA24718.1"  
 /db\_xref="GI:148109"  
 /translation="MKVYQLLAELITSDVRLTIESATQEG"

ORIGIN

Alignment Scores:  
 Pred. No.: 7,02e-81 Length: 2000  
 Score: 854.00 Matches: 165  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 100.0% Indels: 0  
 DB: 15 Gaps: 0

US-10-718-311-4 (1-165) x ECUOBIPLS (1-2000)

Qy	1	MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla	20
Db	380	ATGTCACACCCCGGTTAAACGCACTGCGTGCCTATTTGTTAAAGAGATCCCTGCC	439
Qy	21	LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu	40
Db	440	CTGATCCGCAACTGCTCGACTGGCTGTGTCTGGAGGATTCATGACAAACGTTTGAA	499
Qy	41	GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu	60
Db	500	CAGCAGGGGAAAAACGGTAAGCGTGACGATGATCCCGAAGGGTTGTCCGAGCAGATGAA	559
Qy	61	IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu	80
Db	560	ATCCCCGAAGAACTCCGCTGCTGCCAAGAGTCTCGTTACTGGTTAGCTGAAATTTTG	619
Qy	81	LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu	100
Db	620	TTATGTGCCGATGGTGAACCGTGGCTTCGCCGTGCTACCGTCTCTGTGTCAAGCTTA	679
Qy	101	SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe	120
Db	680	AGCGGGCGGAGCTGGCGTTACAAAAATGGGTAAACCGCCGTAGGACGCTATCTGTT	739
Qy	121	ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly	140
Db	740	ACATCATCGACATTAACCCGGGACTTTATTGAGATAGCCGTGATGCCGGCTGTGGGG	799
Qy	141	ArgArgSerArgLeuArgLeuSerClyLysProLeuLeuLeuThrClnLeuPheLeuPro	160
Db	800	CGAGCTTCCCCCGCTGCGGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTACCG	859
Qy	161	AlaSerProLeuTyr 165	
Db	860	CGCTCACCGTTGTAC 874	
RESULT	15		
ECUBIAC			
LOCUS			
DEFINITION		ECUBIAC 2034 bp DNA linear BCT 05-OCT-1999	
KEYWORDS		E.coli ubiC and ubiA genes for chorismate lyase and 4-hydroxybenzoate octaprenyltransferase.	
ACCESSION		X57434	
VERSION		X57434.1 GI:43233	
KEYWORDS		4-hydroxybenzoate octaprenyl transferase; 4-hydroxybenzoate synthetase; chorismate lyase.	
SOURCE		Escherichia coli	
ORGANISM		Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.	
REFERENCE		1	
AUTHORS		Nishimura, K., Nakahigashi, K. and Inokuchi, H.	
TITLE		Location of the ubiA gene on the physical map of Escherichia coli	
JOURNAL		J. Bacteriol. 174 (17), 5762 (1992)	
PUBMED		1512213	
REFERENCE		2 (bases 1 to 2034)	
AUTHORS		Nishimura, K.	
TITLE		Direct Submission	

JOURNAL Submitted (28-JAN-1991) K. Nishimura, Faculty of Sciences, Kyoto University, Sakyo-ku, Kyoto 606, Japan  
 FEATURES Location/Qualifiers  
 source 1..2034

/organism="Escherichia coli"  
 /mol\_type="genomic DNA"  
 /strain="K12 W3110"  
 /db\_xref="taxon:562"  
 /map="91.7 min"  
 /clone="634"  
 /clone\_lib="Kohara library"

-35\_signal 137..142  
 -10\_signal 159..164  
 gene 291..899

/gene="ubiC"  
 291..899

/gene="ubiC"  
 291..899

/codon\_start=1  
 /transl\_table=11

/product="4-hydroxybenzoate synthetase"  
 /protein\_id="CAA40681.1"

/db\_xref="GI:43234"  
 /db\_xref="GOA:P26602"

/db\_xref="UniProtKB/Swiss-Prot:P26602"  
 /translation="MRLRLFCVLDHLICFTSPVNTFLRYNAFTLCNGFCGMSHPALT

QLRLRYCKEIPALDPQLDLWLLLEDSMTKRFQOQKTVSVTMIREFGEVQNEIPEEL  
 PLLPKESYWLREILLCADGEPWLAGRTVPVPSLPGLALQKLGKTPGLRYLFTSS  
 TLTRDFIEIGRDAGLWGRRLSLGKPLLLTELFPASPLY"

gene 912..1784  
 /gene="ubiA"

912..1784  
 /gene="ubiA"

/codon\_start=1  
 /transl\_table=11

/product="4-hydroxybenzoate octaprenyltransferase"  
 /protein\_id="CAA40682.1"

/db\_xref="GI:43235"  
 /db\_xref="GOA:P26601"

/db\_xref="UniProtKB/Swiss-Prot:P26601"  
 /translation="MWSLTQNLIAFLHRLMRTDKPIGALLLWPTLMALWVATPGVP

QLWLAVVAGVWLMRAAGCVVNDYADRKFQGHVKRTANRPLPSGAVTEKEARALFVV  
 LVLISFLILNTMTILLSTAAALAWVPFMKRYTHLPQVILGAAGWSIPMAFAA  
 VSESVPLSCWLMFLANILMAVAYTOYAMVDRDDDKIGIKSTAILFGQIDKLIIGIL  
 QIVGLALMAIIGELNGWGIYSILVAGALFVYQQLIANREREACFKAFMNNYNG  
 LVLFLGLAMSYWHP"

## ORIGIN

Alignment Scores:  
 Pred. No.: 7,15e-81 Length: 2034  
 Score: 854.00 Matches: 165  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 100.0% Indels: 0  
 DB: 15 Gaps: 0

US-10-718-311-4 (1-165) x ECUBIAC (1-2034)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20  
 Db 402 ATGTACACCCCGCGCTTAACGCAACTGCGTGGCTGCTATTGTAAAGAGATCCCTGCC 461  
 Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40  
 Db 462 CTGGATCCGCAACTGCTCGACTGGCTGTGTGGAGGATTCATGACAAAAACGTTTGGAA 521  
 Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60  
 Db 522 CAGCAGGGGAAAACCGTAAGCGTGACGATGATCCCGAAGGGTTTGTTCGACGAGAAATGAA 581  
 Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80  
 Db 582 ATCCCGGAAGAACTGCCCGCTGTCGCGAAGAGATCTCGTTACTGTTACGTGAAATTTG 641

Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100  
 Db 642 TTATGTGCCGATGGTGAACCGTGGCTTGGCCGTCGTACCGTCGTCTCTGTCACACGTTA 701  
 Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120  
 Db 702 AGCGGGCCGAGCTGGCGTTACAAAANATTGGGTAAACGCCGTTAGGACGCTATCTGTTT 761  
 Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140  
 Db 762 ACATCATCGACATTAAACCGGGACTTTATTGAGATAGCCGTCGATGCCGGCTGTGGGGG 821  
 Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160  
 Db 822 CGAGCTTCCCGCCCTGCGGATTAAAGCGGTAAACCGCTGTGTGCTAACAGAACTGTTTTACCG 881  
 Qy 161 AlaSerProLeuTyr 165  
 Db 882 GCGTCACCGTTGTATC 896

Search completed: May 30, 2006, 05:53:01  
 Job time : 5004.73 secs

GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 29, 2006, 11:01:13 ; Search time 200.955 Seconds  
(without alignments)  
2304.491 Million cell updates/sec

Title: US-10-718-311-4

Perfect score: 854

Sequence: 1 MSHPALTOALRYLCKEIPA.....RLSGKPLLLTELFLPASPLY 165

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 1403666 seqs, 93554401 residues

Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-O=/abs/ABSSWEB spool/US10718311/runat 26052006 164922 13534/app.query.fasta\_1  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=ptc -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-HOST=abs02p -USER=US10718311 @CGN 1 1 615 @runat 26052006 164922 13534  
-NCPU=6 -ICPU=3 -NO MAP -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /EMC\_Celerra\_IDS3/ptodata/2/ina/1 COMB.seq.\*
- 2: /EMC\_Celerra\_IDS3/ptodata/2/ina/5\_COMB.seq.\*
- 3: /EMC\_Celerra\_IDS3/ptodata/2/ina/6A\_COMB.seq.\*
- 4: /EMC\_Celerra\_IDS3/ptodata/2/ina/6B\_COMB.seq.\*
- 5: /EMC\_Celerra\_IDS3/ptodata/2/ina/7\_COMB.seq.\*
- 6: /EMC\_Celerra\_IDS3/ptodata/2/ina/H\_COMB.seq.\*
- 7: /EMC\_Celerra\_IDS3/ptodata/2/ina/PCUTUS\_COMB.seq.\*
- 8: /EMC\_Celerra\_IDS3/ptodata/2/ina/PP\_COMB.seq.\*
- 9: /EMC\_Celerra\_IDS3/ptodata/2/ina/RE\_COMB.seq.\*
- 10: /EMC\_Celerra\_IDS3/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	854	100.0	495	3	US-09-896-866B-12 Sequence 12, Appl
2	854	100.0	495	3	US-09-855-341-3 Sequence 3, Appl
3	854	100.0	684	3	US-09-896-866B-16 Sequence 16, Appl
4	854	100.0	684	3	US-09-855-341-7 Sequence 7, Appl
5	854	100.0	854	3	US-09-064-693A-24 Sequence 24, Appl
6	854	100.0	6641	3	US-09-064-693A-25 Sequence 25, Appl
7	673	78.8	645	3	US-09-489-039A-1234 Sequence 1234, Ap
8	401	47.0	537	3	US-09-543-681A-3122 Sequence 3122, Ap

9	186	21.8	630	3	US-09-252-991A-7442	Sequence 7442, Ap
10	174.5	20.4	519	3	US-09-540-236-1889	Sequence 1889, Ap
11	171.5	20.1	269223	3	US-09-596-002-41	Sequence 41, Appl
12	157.5	18.4	540	3	US-09-328-352-343	Sequence 343, Appl
13	146.5	17.2	1248	3	US-09-252-991A-7361	Sequence 7361, Ap
14	121	14.2	195	3	US-09-543-681A-3086	Sequence 3086, Ap
15	90	10.5	2127	3	US-09-818-780-75	Sequence 75, Appl
16	82	9.7	84839	3	US-09-949-016-15816	Sequence 15816, A
17	82	9.6	601	3	US-09-949-016-69693	Sequence 69693, A
18	82	9.6	807	3	US-09-583-110-1232	Sequence 1232, Ap
19	82	9.6	888	3	US-09-107-433-1721	Sequence 1721, Ap
20	81	9.5	62354	3	US-09-949-016-16188	Sequence 16188, A
21	80	9.4	6463	2	US-08-962-284-3	Sequence 3, Appl
22	80	9.4	68778	3	US-09-949-016-16406	Sequence 16406, A
23	80	9.4	72455	3	US-09-949-016-13793	Sequence 13793, A
24	80	9.4	77535	3	US-09-949-016-14279	Sequence 14279, A
25	80	9.4	77535	3	US-09-949-016-14280	Sequence 14280, A
26	80	9.4	77535	3	US-09-949-016-14281	Sequence 14281, A
27	80	9.4	104475	3	US-09-949-016-12115	Sequence 12115, A
28	80	9.4	111282	3	US-09-754-250-3	Sequence 3, Appl
29	80	9.4	111282	3	US-10-094-989-3	Sequence 3, Appl
30	79.5	9.3	47981	3	US-09-679-279-1	Sequence 1, Appl
31	79.5	9.3	4403765	3	US-09-103-840A-2	Sequence 2, Appl
32	79.5	9.3	4411529	3	US-09-103-840A-1	Sequence 1, Appl
33	79	9.3	1176	3	US-09-252-991A-10640	Sequence 10640, A
34	79	9.3	1581	3	US-09-252-991A-10731	Sequence 10731, A
35	79	9.3	2055	3	US-09-252-991A-10433	Sequence 10433, A
36	79	9.3	5246	3	US-09-799-451-474	Sequence 474, App
37	78.5	9.2	9461	3	US-09-221-017B-513	Sequence 513, App
38	78	9.1	4792	3	US-08-781-891-205	Sequence 205, App
39	78	9.1	4792	3	US-09-618-166-205	Sequence 205, App
40	78	9.1	9173	3	US-09-949-001-30	Sequence 30, Appl
41	78	9.1	9174	2	US-09-949-001-36	Sequence 36, Appl
42	78	9.1	49377	2	US-08-764-233A-1	Sequence 1, Appl
43	77.5	9.1	1090	3	US-09-533-559-3860	Sequence 3860, Ap
44	77.5	9.1	57392	3	US-09-949-016-12070	Sequence 12070, A
45	77.5	9.1	57402	3	US-09-949-016-13293	Sequence 13293, A

ALIGNMENTS

RESULT 1

- US-09-896-866B-12
- Sequence 12, Application US/09896866B
- Patent No. 6642036
- GENERAL INFORMATION:
- APPLICANT: Flint, Dennis
- APPLICANT: Meyer, Knut
- APPLICANT: Viitanen, Paul
- TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugate
- FILE REFERENCE: BC1034 US NA
- CURRENT APPLICATION NUMBER: US/09/896,866B
- PRIOR FILING DATE: 2001-06-29
- PRIOR APPLICATION NUMBER: 60/216,615
- PRIOR FILING DATE: 2000-07-07
- NUMBER OF SEQ ID NOS: 19
- SOFTWARE: Microsoft Office 97
- SEQ ID NO 12
- LENGTH: 495
- TYPE: DNA
- ORGANISM: Escherichia coli
- US-09-896-866B-12

Alignment Scores:			
Pred. No.:	5.04e-98	Length:	495
Score:	854.00	Matches:	165
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	100.0%	Indels:	0
DB:	3	Gaps:	0
US-10-718-311-4 (1-165) x US-09-896-866B-12 (1-495)			

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20  
Db 1 ATGTACACCCCGCGCTTAACCAACTGCGTGGCTGCGCTATTGTTAAAGAGATCCCTGCC 60

Qy 21 LeuAepProGlnLeuLeuAepTrpLeuLeuLeuGluAepSerMetThrLysArgPheGlu 40  
Db 61 CTGGATCCGCAACTGCTGCATGGCTGTTGCTGGAGGATTCTGATGACAAAACGTTTGA 120

Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60  
Db 121 CAGCAGGCAAAACGGTAAGCGTACGATATCCGCAAGGCTTGTGAGCAGAAATGAA 180

Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80  
Db 181 ATCCCGCAAGAACTGCGCTGTCGCAAGAGTCTCGTTACTGTTACGTGAAATTTTG 240

Qy 81 LeuCysAlaAepGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100  
Db 241 TTATGTGCGATGTAACCGTGGCTTCCCGTGGTACCGTCTGTTCTGTCAACGTTA 300

Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120  
Db 301 AGCGGCGGAGTGGCGTTACAAAATTGGTAAACGCCGTTAGGACGCTATCTGTTT 360

Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140  
Db 361 ACATCATCGACATTAACCCGGAGCTTTATTGAGATAGCCGCTGATGCCGGCTGTGGGG 420

Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 160  
Db 421 CGAGTTCGCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAACGAAGTCTGTTTACCG 480

Qy 161 AlaSerProLeuTyr 165  
Db 481 CGGTACCGTGTATC 495

## RESULT 2

US-09-855-341-3  
; Sequence 3, Application US/09855341  
; Patent No. 6683231

GENERAL INFORMATION:  
; APPLICANT: VIITANEN, PAUL V.

APPLICANT: MEYER, KNU

APPLICANT: VAN DYK, DREW

TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID

TITLE OF INVENTION: IN GREEN PLANTS

FILE REFERENCE: BC1015 US NA

CURRENT APPLICATION NUMBER: US/09/855,341

CURRENT FILING DATE: 2001-05-15

NUMBER OF SEQ ID NOS: 16

SOFTWARE: MICROSOFT OFFICE 97

SEQ ID NO 3

LENGTH: 495

TYPE: DNA

ORGANISM: Unknown Organism

FEATURE:

OTHER INFORMATION: Description of Unknown Organism:E. coli

US-09-855-341-3

Alignment Scores:  
Pred. No.: 5,04e-98 Length: 495  
Score: 854.00 Matches: 165  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 3 Gaps: 0

US-10-718-311-4 (1-165) x US-09-855-341-3 (1-495)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20  
Db 1 ATGTACACCCCGCGCTTAACCAACTGCGTGGCTGCGCTATTGTTAAAGAGATCCCTGCC 60

Qy 21 LeuAepProGlnLeuLeuAepTrpLeuLeuLeuGluAepSerMetThrLysArgPheGlu 40  
Db 61 CTGGATCCGCAACTGCTGCATGGCTGTTGCTGGAGGATTCTCATGACAAAACGTTTGA 120

Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60  
Db 121 CAGCAGGCAAAACGGTAAGCGTACGATATCCGCAAGGCTTGTGAGCAGAAATGAA 180

Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80  
Db 181 ATCCCGCAAGAACTGCGCTGTCGCAAGAGTCTCGTTACTGTTACGTGAAATTTTG 240

Qy 81 LeuCysAlaAepGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100  
Db 241 TTATGTGCGATGTAACCGTGGCTTCCCGTGGTACCGTCTGTTCTGTCAACGTTA 300

Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120  
Db 301 AGCGGCGGAGTGGCGTTACAAAATTGGTAAACGCCGTTAGGACGCTATCTGTTT 360

Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140  
Db 361 ACATCATCGACATTAACCCGGAGCTTTATTGAGATAGCCGCTGATGCCGGCTGTGGGG 420

Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 160  
Db 421 CGAGTTCGCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAACGAAGTCTGTTTACCG 480

Qy 161 AlaSerProLeuTyr 165  
Db 481 CGGTACCGTGTATC 495

## RESULT 3

US-09-896-866B-16  
; Sequence 16, Application US/09896866B  
; Patent No. 6642036

GENERAL INFORMATION:

APPLICANT: Flint, Dennis

APPLICANT: Meyer, Knut

APPLICANT: Viitanen, Paul

TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugates

FILE REFERENCE: BC1034 US NA

CURRENT APPLICATION NUMBER: US/09/896,866B

CURRENT FILING DATE: 2001-06-29

PRIOR APPLICATION NUMBER: 60/216,615

PRIOR FILING DATE: 2000-07-07

NUMBER OF SEQ ID NOS: 19

SOFTWARE: Microsoft Office 97

SEQ ID NO 16

LENGTH: 684

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

NAME/KEY: misc feature

OTHER INFORMATION: open reading frame of the chloroplast-targeted CPL fusion protein

US-09-896-866B-16

Alignment Scores:  
Pred. No.: 8,29e-98 Length: 684  
Score: 854.00 Matches: 165  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 3 Gaps: 0

US-10-718-311-4 (1-165) x US-09-896-866B-16 (1-684)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20  
Db 187 ATGTACACCCCGCGCTTAACCAACTGCGTGGCTGCGCTATTGTTAAAGAGATCCCTGCC 246

QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40  
Db 247 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGATTCCATGACAAAACGTTTTTGA 306  
QY 41 GlnGlnGlyThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60  
Db 307 CAGCAGGGGAAAAACGGTAGCGTAGATCCGCGAAGAGGTTCCTGAGCAGAGATGAA 366  
QY 61 IleProGluGluLeuProLeuProLysGluSerArgTyTrpLeuArgGluIleLeu 80  
Db 367 ATCCCGGAGAACTGCGCTGTGCGCGAAGAGTCTCGTTACTGTTAGTGAATTTTG 426  
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100  
Db 427 TTATGTGCCGATGGTGAACCGTGGCTTGCCTGCTGCTGCTCAACGTTA 486  
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyLeuPhe 120  
Db 487 AGCGGGCCGAGCTGGCGTTACAAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 546  
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140  
Db 547 ACATCATCGACATTAAACCGGACTTATTAGATAGGCGGTGATGCCGGCTGTGGGG 606  
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160  
Db 607 CGACGTTCCCGCTGCGATTAGCGGTAAACCGCTGTTCTTAACAGAACTGTTTTTACCG 666  
QY 161 AlaSerProLeuTytr 165  
Db 667 GCGTCACCGTTGTAC 681

## RESULT 4

US-09-855-341-7  
; Sequence 7, Application US/09855341  
; Patent No. 668231  
; GENERAL INFORMATION:  
; APPLICANT: VIITANEN, PAUL V.  
; APPLICANT: MEYER, KNUT  
; APPLICANT: VAN DYK, DREW  
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID  
; FILE REFERENCE: IN GREEN PLANTS  
; CURRENT APPLICATION NUMBER: US/09/855,341  
; CURRENT FILING DATE: 2001-05-15  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: MICROSOFT OFFICE 97  
; SEQ ID NO 7  
; LENGTH: 684  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:synthetic CPL  
US-09-855-341-7

## Alignment Scores:

Pred. No.:	8,298-98	Length:	684
Score:	854.00	Matches:	165
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	100.0%	Indels:	0
DB:	3	Gaps:	0

US-10-718-311-4 (1-165) x US-09-855-341-7 (1-684)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyCysLysGluIleProLa 20  
Db 187 ATGTCACACCCCGCTTAACGCAACTGCGTGGCTATTTGTAAGAGATCCCTGCC 246  
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40  
Db 247 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGATTCCATGACAAAACGTTTTTGA 306

QY 41 GlnGlnGlyThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60  
Db 307 CAGCAGGGGAAAAACGGTAGCGTAGATCCGCGAAGAGGTTCCTGAGCAGAGATGAA 366  
QY 61 IleProGluGluLeuProLeuProLysGluSerArgTyTrpLeuArgGluIleLeu 80  
Db 367 ATCCCGGAGAACTGCGCTGTGCGCGAAGAGTCTCGTTACTGTTAGTGAATTTTG 426  
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100  
Db 427 TTATGTGCCGATGGTGAACCGTGGCTTGCCTGCTGCTGCTCAACGTTA 486  
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyLeuPhe 120  
Db 487 AGCGGGCCGAGCTGGCGTTACAAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 546  
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140  
Db 547 ACATCATCGACATTAAACCGGACTTATTAGATAGGCGGTGATGCCGGCTGTGGGG 606  
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160  
Db 607 CGACGTTCCCGCTGCGATTAGCGGTAAACCGCTGTTCTTAACAGAACTGTTTTTACCG 666  
QY 161 AlaSerProLeuTytr 165  
Db 667 GCGTCACCGTTGTAC 681

## RESULT 5

US-09-064-693A-24  
; Sequence 24, Application US/09064693A  
; Patent No. 6210937  
; GENERAL INFORMATION:  
; APPLICANT: Ward, Thomas E.  
; TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY  
; TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: W. Gary Goodson  
; ADDRESSEE: INEEL--Lockheed Martin Idaho  
; ADDRESSEE: Technologies Co.  
; STREET: P.O. Box 1625  
; CITY: Idaho Falls  
; STATE: Idaho  
; COUNTRY: USA  
; ZIP: 83415-3810  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: Toshiba Satellite Pro T2150CDS  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: Word Perfect 7.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/064,693A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: W. Gary Goodson  
; REGISTRATION NUMBER: 22,387  
; REFERENCE/DOCKET NUMBER: LIT-PI-296  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (208)526-9469  
; TELEFAX: (208)526-8339  
; INFORMATION FOR SEQ ID NO: 24:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 854 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-09-064-693A-24



```
Alignment Scores:
Pred. No.: 1,17e-97 Length: 854
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0

US-10-718-311-4 (1-165) x US-09-064-693A-24 (1-854)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 20
Db 321 ATGTCACACCCCGGTTAACGCACTGCTGCGCTATTTAAAGAGATCCCTGCC 380
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 381 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGAA 440
Qy 41 GlnGlnGlyLysThrValSerValThrMetLeuArgGluGlyPheValGluGlnAsnGlu 60
Db 441 CAGCAGGGGAAACCGTTAAGCGTGACGATATCCCGAAGGGTTTGTGAGCAGAAATGAA 500
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeu 80
Db 501 ATCCCGGAAGAACTGCCCTGCTGCCGAAAGAGTCTGTTACTGTTACGTGAAATTTTG 560
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 561 TTATGTCCGATGGTGAACCGTGGCTTCCGCTGCTACCGTCTCTGTGTCACAGTTA 620
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 621 AGCGGGCCGAGCTGGCGTTACAAAAATGGGTAACCGCTGTTAGGACGCTATCTGTTTC 680
Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluLeuGlyArgAspAlaGlyLeuTrpGly 140
Db 681 ACATCATCGACATTAACCCGGGACTTTATTAGATAGCCGCTGATGCCGCTGTGGGG 740
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 741 CGAGCTTCCCGCTCGGATTAAGCGGTAACCGCTGTTGCTTAACAGAACTGTTTTTACCG 800
Qy 161 AlaSerProLeuTyr 165
Db 801 CGGTACCGTTGTAC 815

RESULT 6
US-09-064-693A-25
; Sequence 25, Application US/09064693A
; Patent No. 6210937
; GENERAL INFORMATION:
; APPLICANT: Ward, Thomas E.
; TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY
; TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION
; TITLE OF INVENTION: OF A SPECIFIC PLASTICS PRECURSOR
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W. Gary Goodson
; ADDRESSEE: INEEL--Lockheed Martin Idaho
; ADDRESSEE: Technologies Co.
; STREET: P.O. Box 1625
; CITY: Idaho Falls
; STATE: Idaho
; COUNTRY: USA
; ZIP: 83415-3810
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: Toshiba Satellite Pro T2150CDS
; OPERATING SYSTEM: Windows95
; SOFTWARE: Word Perfect 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/064,693A

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: W. Gary Goodson
; REGISTRATION NUMBER: 22,387
; REFERENCE/DOCKET NUMBER: LIT-PI-296
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (208) 526-9469
; TELEFAX: (208) 526-8339
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6641 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-064-693A-25

Alignment Scores:
Pred. No.: 2,74e-96 Length: 6641
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0

US-10-718-311-4 (1-165) x US-09-064-693A-25 (1-6641)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 20
Db 321 ATGTCACACCCCGGTTAACGCACTGCTGCGCTATTTAAAGAGATCCCTGCC 380
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 381 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGAA 440
Qy 41 GlnGlnGlyLysThrValSerValThrMetLeuArgGluGlyPheValGluGlnAsnGlu 60
Db 441 CAGCAGGGGAAACCGTTAAGCGTGACGATATCCCGAAGGGTTTGTGAGCAGAAATGAA 500
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeu 80
Db 501 ATCCCGGAAGAACTGCCCTGCTGCCGAAAGAGTCTGTTACTGTTACGTGAAATTTTG 560
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 561 TTATGTCCGATGGTGAACCGTGGCTTCCGCTGCTACCGTCTCTGTGTCACAGTTA 620
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 621 AGCGGGCCGAGCTGGCGTTACAAAAATGGGTAACCGCTGTTAGGACGCTATCTGTTTC 680
Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluLeuGlyArgAspAlaGlyLeuTrpGly 140
Db 681 ACATCATCGACATTAACCCGGGACTTTATTAGATAGCCGCTGATGCCGCTGTGGGG 740
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 741 CGAGCTTCCCGCTCGGATTAAGCGGTAACCGCTGTTGCTTAACAGAACTGTTTTTACCG 800
Qy 161 AlaSerProLeuTyr 165
Db 801 CGGTACCGTTGTAC 815

RESULT 7
US-09-489-039A-1234
; Sequence 1234, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
```



;; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
;; FILE REFERENCE: 2709.2004001  
;; CURRENT APPLICATION NUMBER: US/09/489,039A  
;; CURRENT FILING DATE: 2000-01-27  
;; PRIOR APPLICATION NUMBER: US 60/117,747  
;; PRIOR FILING DATE: 1999-01-29  
;; NUMBER OF SEQ ID NOS: 14342  
;; SEQ ID NO 1234  
;; LENGTH: 645  
;; TYPE: DNA  
;; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-1234

Alignment Scores:  
Pred. No.: 4,54e-75 Length: 645  
Score: 673.00 Matches: 129  
Percent Similarity: 87.9% Conservative: 16  
Best Local Similarity: 78.2% Mismatches: 20  
Query Match: 78.8% Indels: 0  
DB: 3 Gaps: 0

US-10-718-311-4 (1-165) x US-09-489-039A-1234 (1-645)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20  
Db 148 ATGTCCCATCTCGCGCTTACGGGACTGCGTGGCGCTATTTTCCCGTTATGCCATCC 207  
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40  
Db 208 CTGCGCGCCCGCTAGTCACTGGCTGCTGCTGGAGGACTCAATGACCCAGCGCTTGAA 267  
QY 41 GlnGlnGlyThrValSerValThrMetIleArgGluGlyPheValGluGlnAenGlu 60  
Db 268 CAACAAGGAACAGGTACCGTACCGTGGTTACGAAGGTATATCGCGCGTACGCG 327  
QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80  
Db 328 CTGACTGACGAACGCGCGCTGCTGCCGACGAGCGCGCTACTGGCTGGCGGAGATTATA 387  
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100  
Db 388 CTCAATGCGATGGGAGCCCTGGCTGGCGCGGACGCTGGCGCGGAGTCGACGCTG 447  
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120  
Db 448 TGTGTCGCGAGCTGGCGCTACAGAGCTCGGGCAACTCCGCTGGCGGATACCTGTT 507  
QY 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140  
Db 508 ACGTCGTCGACATTAAACCGCGGATTTTATTGAAATTGGTTCGCGATGCGCGCTGTGGGG 567  
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160  
Db 568 CCGTCGTTCCCGCTGGCGGTGAGCGGCAAAACCCCTGCTGCTACCGAGCTTTTGTGCT 627  
QY 161 AlaSerProLeuTyr 165  
Db 628 GCGTCGCGCTGTAC 642

## RESULT 8

US-09-543-681A-3122  
; Sequence 3122, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON

;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
;; FILE REFERENCE: 2709.1002-001  
;; CURRENT APPLICATION NUMBER: US/09/543,681A  
;; CURRENT FILING DATE: 2000-04-05  
;; PRIOR APPLICATION NUMBER: US 60/128,706  
;; PRIOR FILING DATE: 1999-04-09  
;; NUMBER OF SEQ ID NOS: 8344

;; SEQ ID NO 3122  
;; LENGTH: 537  
;; TYPE: DNA  
;; ORGANISM: Proteus mirabilis  
US-09-543-681A-3122

Alignment Scores:  
Pred. No.: 5.8e-41 Length: 537  
Score: 401.00 Matches: 81  
Percent Similarity: 68.6% Conservative: 15  
Best Local Similarity: 57.9% Mismatches: 44  
Query Match: 47.0% Indels: 0  
DB: 3 Gaps: 0

US-10-718-311-4 (1-165) x US-09-543-681A-3122 (1-537)

QY 26 LeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGluGlnGlnGlyLysThr 45  
Db 109 TTAAGTTGGCTGCTGGAATTAGGTTCATGACACGACGTTTTCAGCAACATTCGCATCAA 168  
QY 46 ValSerValThrMetIleArgGluGlyPheValGluGlnAsnGluIleProGluGluLeu 65  
Db 169 GTGACGGTAATGCCATATCAAGAAGGTTTATTGAATATATCGAGCTGCTGATGAACAA 228  
QY 66 ProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeuLeuCysAlaAspGly 85  
Db 229 AGTGTTTACCTTATAGCCGCGCTATTGGCTAAGAGAATTCGCTTTTGTGGGGAATAT 288  
QY 86 GluProTrpLeuAlaGlyArgThrValValProValSerThrLeuSerGlyProGluLeu 105  
Db 289 GTACCTTGGTTATTAGGCGCAACATTAGTGCAGGAGACATTACCGGTGAAGATCGC 348  
QY 106 AlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrTrpLeuPheThrSerThrLeu 125  
Db 349 CAATTAGTGAATTGGTACGTCGTCATAGGACGTTATCTGTTCAGAGAAACAACCTTA 408  
QY 126 ThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGlyArgArgSerArgLeu 145  
Db 409 ACTCGTGATTTTATTCATATTGGGCAACAAATGGACATTGGTACGTCGTTCCCGTTTC 468  
QY 146 ArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuProAlaSerProLeuTyr 165  
Db 469 CAGCTTTCAGATAAACCTTTATTATTACTAGAGTGTGTTTTCGCTGCATCACCTGTATAT 528

## RESULT 9

US-09-252-991A-7442  
; Sequence 7442, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.

;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
;; FILE REFERENCE: 107196.136  
;; CURRENT APPLICATION NUMBER: US/09/252,991A  
;; CURRENT FILING DATE: 1999-02-18  
;; PRIOR APPLICATION NUMBER: US 60/074,788  
;; PRIOR FILING DATE: 1998-02-18  
;; PRIOR APPLICATION NUMBER: US 60/094,190  
;; PRIOR FILING DATE: 1998-07-27  
;; NUMBER OF SEQ ID NOS: 33142  
;; SEQ ID NO 7442  
;; LENGTH: 630  
;; TYPE: DNA  
;; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7442

Alignment Scores:  
Pred. No.: 8.44e-14 Length: 630  
Score: 186.00 Matches: 50  
Percent Similarity: 43.0% Conservative: 21  
Best Local Similarity: 30.3% Mismatches: 64  
Query Match: 21.8% Indels: 30  
DB: 3 Gaps: 3

```
US-10-718-311-4 (1-165) x US-09-252-991A-7442 (1-630)
Qy 17 GluileProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThr 36
Db 127 CAACGTTCAGAGCGTCTCCCGCGCTCTCGACTGGCTGTCGACGAGGGGTGCGTGACC 186
Qy 37 LysArgPheGluGlnGlnGlyThrValSerValThrMetIleArgGluGlyPheVal 56
Db 187 CGGCGC-----CTGACCGCCCTGGCGCGCGCGCGTTC 219
Qy 57 GluGlnAanGluIleProGluGluLeuProLeuLeu----- 68
Db 220 CGCGTCGAACCCCTCTCGAAGGCTGGCGAGCCCTGCGCGAGCGAGCAATGCCAGGGGCTC 279
Qy 69 -----ProLysGluSerArgTyTrpLeuArgGluIleLeuLeuCysAlaAspGlyGlu 86
Db 280 GACGTCCCTACCGGAGTAGCGGTGGCTGGCGAGGTCACCTGCATGGCCACGACCGT 339
Qy 87 ProTrpLeuAlaGlyArgThrValProValSerThrLeuSerGlyProGluLeuAla 106
Db 340 CCCTGGGTGTTGCGCGCGAGGTGGCGGCGCGAGCCCTGGAAGGCTCGGGCTTCGAC 399
Qy 107 LeuGlnLysLeuGlyLysThrProLeuGlyArgTyTrpLeuPheThrSerThrLeuThr 126
Db 400 CTGGCGCTGCTCGGACCCCGCTGCTGGCGGAGTTGCTTCAGCGACGAGCGCTTCGAG 459
Qy 127 ArgAspPheIleGluIle-----GlyArgAsp 135
Db 460 CGCGGGCCATCGAAGTCTCGCGCTATCCGGCGCGGTCTGCCCGCGAGGTCCGCGCC 519
Qy 136 AlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 155
Db 520 GAGGCTCTCTGGGCGCGTCTGCTACGCTTTTCCCGCGCGCGCTCGGGGTGCTGTTGGCG 579
Qy 156 GluLeuPheLeuPro 160
Db 580 GAGGTGTACTACCG 594

RESULT 10
US-09-540-236-1889
; Sequence 1889, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 1889
; TYPE: DNA
; ORGANISM: M.catarrhalis
US-09-540-236-1889

Alignment Scores:
Pred. No.: 1,75e-12 Length: 519
Score: 174.50 Matches: 48
Percent Similarity: 49.4% Conservative: 36
Best Local Similarity: 28.2% Mismatches: 57
Query Match: 20.4% Indels: 29
DB: 3 Gaps: 7

US-10-718-311-4 (1-165) x US-09-540-236-1889 (1-519)
Qy 15 CysLysGluIleProAlaLeuAsp----- 22
Db 4 TGTAACATTGACCGCCCAATAGATGATTTTCATGCCAATGGATGATATAGACCAC 63
Qy 23 -----ProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 23 -----ProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40

US-10-718-311-4 (1-165) x US-09-596-002-41 (1-269223)
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 16824 CTGCTCCAAACTTGGCGGATTTGTATTCTACTCAAGGTCGTACGGCAGGATTAGAA 16883
Qy 41 GlnGlnGly---LysThrValSerValThrMetIleArgGly----- 54
Db 16884 GCACACTCAAAAAAACCCTTAAAGGTTGAAATCATCATCAAGGCTATCAGCCACTCACC 16943
```

Qy	55	PheValGluGlnIleAsnGluIleProGluLeuLeuProLeuLeuProLysGluSerArgTyr	74
Db	16944	TTTTCCCAAAAATCACAACTG-----GGACTGCC---CTAATCGCCACAAATGGCT	16994
Qy	75	TripLeuArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrVal	94
Db	16995	TGGGTAGCAGTGTCAGATTATACGGTGATGCGGCGCTTGGTGCTGCCAAAAAGTNT	17054
Qy	95	ValProValSerThrLeuSerGlyProGluLeuLeuAlaGluLeuGlyLysThrPro	114
Db	17055	TTTCCTTTGTCACTCACTGCAGGTTTCATTAGCACGCTATAAAGTACTGGGTCAACGCC	17114
Qy	115	LeuClYArgTyrLeuPheThrSerSer-----ThrLeuThrArgAspPheIleGlu	131
Db	17115	ATCGGATATGTTTGTTCATAAAAAAGCAACAACACTGCCAATTAAGCGTCATTTTAT---	17171
Qy	132	IleGlyArgAspAlaGlyLeuTrpGlyArgArgSerArgLeuArgLeuSerGlyLysPro	151
Db	17172	-----CGGTATGATGGTCAATTTGGTCGGAACAACAGCTTATCATATCAATGGGGCTAAA	17225
Qy	152	LeuLeuLeuThrGluLeuPheLeuProAla	161
Db	17226	ATCTTGATTGATGAGCTATTTTGGCAGCT	17255

## RESULT 12

```

US-09-328-352-343
; Sequence 343, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND
; TITLE OF INVENTION: BAUMANNII FOR DI-
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8352
; SEQ ID NO 343
; LENGTH: 540
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-343

```

Alignment Scores:		
Pred. No.:	2,57e-10	540
Score:	157.50	44
Percent Similarity:	45.9%	Conservative: 24
Best Local Similarity:	49.7%	Mismatches: 77
Query Match:	18.4%	Indels: 3
DB:	3	Gaps: 3

US-10-718-311-4 (1-165) x US-09-328-352-343 (1-540)

Qy		16	LysGluIleProAlaLeuAspP	ProGlnLeuLeuAspTrp	LeuLeuLeuGluA	SerMet	35
			:::::	:::::	:::::	:::::	
Db		55	AAGCAGAAAGAAGACTT	TAAATCGAATTGA	AAACATGGTTGT	TGCCTTCGTCG	114
			:::::	:::::	:::::	:::::	
Qy		36	ThrLysArgPhe---	GluGlnGlnGlySerVal	SerValThrMet	IleArgGluGly	54
			:::::	:::::	:::::	:::::	
Db		115	ACAACAACA	CTCTGAGCTGGTG	TGGGAAGTTCA	GCCTTTCAAAGAACA	174
			:::::	:::::	:::::	:::::	
Qy		55	PheValGluGlnAsnGlu	IleProGluLeuLeuPro	LeuLeuProLysGlu	SerArgTyr	74
			:::::	:::::	:::::	:::::	
Db		175	TTCCAGCGTTAACT	TTTCCGACAGTCAA	TGGATGAACATGCC	CCCCATACTCAC	234
			:::::	:::::	:::::	:::::	
Qy		75	TrpLeuArgGluIleLeu	LeuLeuCys--AlaAsp	GlyGluProTrpLeu	AlaGlyArgThr	93
			:::::	:::::	:::::	:::::	
Db		235	TGGTGAGGNAACCT	ATTATATATGCAGT	GATGTAGAGCCTTGG	TGGTAAGCMAAAGT	294
			:::::	:::::	:::::	:::::	
Qy		94	ValValProValSerThr	LeuSerGlyProGluLeu	AlaLeuGlnLysLeu	GlyLysThr	113
			:::::	:::::	:::::	:::::	
Db		295	ATTTTTCCAATCAA	AGTTTACAAAAA	AGCCCGCTATATTT	TACGATATTTGGT	354
			:::::	:::::	:::::	:::::	

Qy	114	ProLeuGlyArgTyrLeupheThrSerSerThrLeuThrArgAspPheLeuGluLeGly	133
Db	355	CGATAGGTCTTTTTTTTTTCAAGACACACACCACTTTGTGATCGCCGGGTATTTCGT	414
Qy	134	ArgAspAlaGlyLeuTrpGlyArgGserArgLeuArgLeuSerGlyLysProLeuLeu	153
Db	415	TTACCAGAAGGC--TGACGGCAGCAAAAGTTGTTACTTGGCATGATGTAATAATTATT	471
Qy	154	LeuThrGluLeuPheLeuProAla	161
Db	472	GTTCAAGAAACATTTCTACCGGCT	495

RESULT 13

```

US-09-252-991A-7361
; Sequence 7361, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7361
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7361

```

**Alignment Scores:**

Pred. No.:	2.26e-08	Length:	1248
Score:	146.50	Matches:	35
Percent Similarity:	46.4%	Conservative:	10
Best Local Similarity:	36.1%	Mismatches:	41
Query Match:	17.2%	Indels:	11
DB:	3	Gaps:	1

US-10-718-311-4 (1-165) x US-09-252-991A-7361 (1-1248)

Qy	75	TrpLeuArgGluLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrVal	94
Db	5	TGGGTCCGCGAGGTCTACTCGATGGCCACGACCGTCTGGGTCTGCCCGCAGCGTG	64
Qy	95	ValProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrPro	114
Db	65	CGCGCGGCAGCGCCTGGAGGGCTCGGGCTTCGACCTGGCGCTGTCGGCACCCGCTCG	124
Qy	115	LeuGlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluLeu	132
Db	125	CTGGCGAGTGTGCTGTTTCAGCGACAGCGCCTTCGAGCGCGGCCCATCGAAGTCTCGCGC	184
Qy	133	-----GlyArgAspAlaGlyLeuTrpGlyArgArgSer	143
Db	185	TATCCGGCGCGGTCTGCCCGCGAGGTCCGCGCGAGGGTCTCTGGGGCGGTGCTCTCA	244
Qy	144	ArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro	160
Db	245	CGGTTTTCCCGCGCGCGCTCGGGGTGCTGTGGCGGAGGTGTACTACTCG	295

**RESULT 14**

US-09-543-681A-3086/c  
Sequence 3086, Application US/09543681A  
Patent No. 6605709  
GENERAL INFORMATION:  
APPLICANT: GARY BRETON  
TITLE OF INVENTION: NUCLEIC ACID AND A  
TITLE OF INVENTION: DIAGNOSTICS AND T  
FILE REFERENCE: 2709.1002-001

```

: APPLICANT: GARY BRETON
:
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
:
: TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
:
: FILE REFERENCE: 2709.1002-001

```



GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 30, 2006, 03:03:17 ; Search time 1135.79 Seconds

(without alignments)  
2677.596 Million cell updates/sec

Title: US-10-718-311-4

Perfect score: 854

Sequence: 1 MSHPALQRLALRYCKEIPA.....RLSGKPLLTLELFPASPLY 165

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 18892170 seqs, 6143817638 residues

Total number of hits satisfying chosen parameters: 37784340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/abss/ABSSWEB/spool/US10718311/runat\_26052006\_164933\_13735/app\_query.fasta\_1  
-DB=PublishedApplications\_NA\_Main -QFMT=fastap -SUFFIX=rnpbm -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -HOST=abss05h  
-USER=US10718311 @CGN 1 1 2326 @runat\_26052006\_164933\_13735 -NCPU=6 -ICPU=3  
-NO\_WMAP -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DBLEXT=7

Database : Published Applications NA\_Main:

1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
7: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
8: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
9: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*  
10: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*  
11: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10F\_PUBCOMB.seq.\*  
12: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10G\_PUBCOMB.seq.\*  
13: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10H\_PUBCOMB.seq.\*  
14: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10I\_PUBCOMB.seq.\*  
15: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10J\_PUBCOMB.seq.\*  
16: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10K\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	854	100.0	495	3	US-09-855-341-3
Sequence 3, Appli					

2	854	100.0	495	3	US-09-896-866B-12	Sequence 12, Appl
3	854	100.0	495	7	US-10-359-369-37	Sequence 37, Appl
4	854	100.0	495	7	US-10-699-050-12	Sequence 12, Appl
5	854	100.0	495	8	US-10-718-311-3	Sequence 3, Appli
6	854	100.0	498	3	US-10-462-162-29	Sequence 29, Appl
7	854	100.0	684	3	US-09-855-341-7	Sequence 7, Appli
8	854	100.0	684	3	US-09-896-866B-16	Sequence 16, Appl
9	854	100.0	684	7	US-10-359-369-41	Sequence 41, Appl
10	854	100.0	684	8	US-10-699-050-16	Sequence 16, Appl
11	854	100.0	684	8	US-10-718-311-7	Sequence 7, Appli
12	854	100.0	684	9	US-10-462-162-33	Sequence 33, Appl
13	854	100.0	3452	9	US-10-462-162-47	Sequence 47, Appl
14	854	98.7	1971	9	US-10-462-162-54	Sequence 54, Appl
15	171.5	20.1	1207	10	US-10-450-763-26139	Sequence 26139, A
16	171.5	20.1	269233	8	US-10-672-787-41	Sequence 41, Appl
17	97.5	11.4	475	8	US-10-275-026A-113	Sequence 113, App
18	91	10.7	606	4	US-09-925-065A-789463	Sequence 789463,
19	91	10.7	606	5	US-09-925-065A-789463	Sequence 789463,
20	90	10.5	2127	9	US-10-746-167-75	Sequence 75, Appl
21	90	10.5	9133	10	US-10-915-740A-277	Sequence 277, App
22	90	10.5	2242716	10	US-10-915-740A-1068	Sequence 1068, Ap
23	88	10.3	42811	11	US-10-330-773-925	Sequence 925, App
24	87	10.2	59589	6	US-10-087-192-1315	Sequence 1315, Ap
25	86	10.1	2892	16	US-11-183-136-15	Sequence 15, Appl
26	85.5	10.0	522	10	US-10-467-657-2855	Sequence 2855, Ap
27	84.5	9.9	1780	10	US-10-487-078-92	Sequence 92, Appl
28	84.5	9.9	111331	10	US-10-461-862-101	Sequence 101, App
29	83.5	9.8	1565	10	US-10-750-185-3285	Sequence 3285, A
30	83.5	9.8	1565	10	US-10-750-623-3285	Sequence 3285, A
31	83.5	9.8	10710	8	US-10-672-396-4	Sequence 4, Appli
32	83	9.7	608	4	US-09-925-065A-772942	Sequence 772942,
33	83	9.7	608	5	US-09-925-065A-772942	Sequence 772942,
34	82.5	9.7	689	13	US-11-097-143-18275	Sequence 18275, A
35	82.5	9.7	122336	8	US-10-322-281-362	Sequence 362, App
36	82	9.6	444	12	US-10-301-480-289318	Sequence 289318,
37	82	9.6	444	12	US-10-301-480-902727	Sequence 902727,
38	82	9.6	452	4	US-09-925-065A-200730	Sequence 200730,
39	82	9.6	452	5	US-09-925-065A-200730	Sequence 200730,
40	82	9.6	888	10	US-10-617-320-1721	Sequence 1721, Ap
41	81.5	9.5	6369	8	US-10-437-963-11402	Sequence 11402, A
42	81	9.5	435	4	US-09-925-065A-200729	Sequence 200729,
43	81	9.5	435	5	US-09-925-065A-200729	Sequence 200729,
44	81	9.5	439	12	US-10-301-480-289317	Sequence 289317,
45	81	9.5	439	12	US-10-301-480-902726	Sequence 902726,

ALIGNMENTS

RESULT 1

US-09-855-341-3  
; Sequence 3, Application US/09855341  
; Patent No. US2002002715A1  
; GENERAL INFORMATION:  
; APPLICANT: VIITANEN, PAUL V.  
; APPLICANT: MEYER, KNUT  
; APPLICANT: VAN DYK, DREW  
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID  
; TITLE OF INVENTION: IN GREEN PLANTS  
; FILE REFERENCE: BC1015 US NA  
; CURRENT APPLICATION NUMBER: US/09/855,341  
; CURRENT FILING DATE: 2001-05-15  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: MICROSOFT OFFICE 97  
; SEQ ID NO 3  
; LENGTH: 495  
; TYPE: DNA  
; ORGANISM: Unknown Organism  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism:E. coli  
US-09-855-341-3

Alignment Scores:

Pred. No.: 1.91e-102 Length: 495

Score: 854.00 Matches: 165  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 3 Gaps: 0

US-10-718-311-4 (1-165) x US-09-855-341-3 (1-495)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 20  
Db 1 ATGTCACACCCCGCGTTAACGCAACTGCGTGGCTGCTATTGTAAAGAGATCCCTGCC 60

Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40  
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTGTGGAGGATTCATGACAAAAACGTTTGA 120

Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAsnGlu 60  
Db 121 CAGCAGGGAAGAAACGGTAAGCGTGACGATGATCCGGAAGGGTTGTGACAGAGATGAA 180

Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeuLeu 80  
Db 181 ATCCCGGAAGAACTGCCGCTGCTGCCGAAGAGTCTCGTTACTGTTACGTGAAATTTG 240

Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100  
Db 241 TTATGTGCGGATGGTAACCGTGGCTGCTCCCGTACCGTCTGTTCTGTCAACGTTA 300

Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120  
Db 301 AGCGGGCCGAGTGGCGTTACAAAAATTTGGTAAGCCGTTAGGACGCTATCTGTTTC 360

Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluLeuGlyArgAspAlaGlyLeuTrpGly 140  
Db 361 ACATCATCGACATTAACCCCGGACTTTATTGAGATAGCCGCTGATGCCGGCTGTGGGG 420

Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160  
Db 421 CGACGTTCCCGCTGCGATTAAAGCGGTAACCGCTGTTGCTTAACAGAACTGTTTTACCG 480

Qy 161 AlaSerProLeuTyr 165  
Db 481 CGCTCACCGTTGTATC 495

RESULT 2  
US-09-896-866B-12  
; Sequence 12, Application US/09896866B  
; Patent No. US20020151002A1  
; GENERAL INFORMATION:  
; APPLICANT: Flint, Dennis  
; APPLICANT: Meyer, Knut  
; APPLICANT: Viitanen, Paul  
; TITLE OF INVENTION: Sinapoylgucose:Malate Sinapoyltransferase Form Malate Conjugates  
; FILE REFERENCE: BC1034 US NA  
; CURRENT APPLICATION NUMBER: US/09/896.866B  
; CURRENT FILING DATE: 2001-06-29  
; PRIOR FILING DATE: 60/216,615  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 12  
; LENGTH: 495  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
US-09-896-866B-12

Alignment Scores:  
Pred. No.: 1.91e-102 Length: 495  
Score: 854.00 Matches: 165  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0

DB: 3 Gaps: 0

US-10-718-311-4 (1-165) x US-09-896-866B-12 (1-495)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 20  
Db 1 ATGTCACACCCCGCGTTAACGCAACTGCGTGGCTGCTATTGTAAAGAGATCCCTGCC 60

Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40  
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTGTGGAGGATTCATGACAAAAACGTTTGA 120

Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAsnGlu 60  
Db 121 CAGCAGGGAAGAAACGGTAAGCGTGACGATGATCCGGAAGGGTTGTGACAGAGATGAA 180

Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeuLeu 80  
Db 181 ATCCCGGAAGAACTGCCGCTGCTGCCGAAGAGTCTCGTTACTGTTACGTGAAATTTG 240

Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100  
Db 241 TTATGTGCGGATGGTAACCGTGGCTGCTCCCGTACCGTCTGTTCTGTCAACGTTA 300

Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120  
Db 301 AGCGGGCCGAGTGGCGTTACAAAAATTTGGTAAGCCGTTAGGACGCTATCTGTTTC 360

Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluLeuGlyArgAspAlaGlyLeuTrpGly 140  
Db 361 ACATCATCGACATTAACCCCGGACTTTATTGAGATAGCCGCTGATGCCGGCTGTGGGG 420

Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160  
Db 421 CGACGTTCCCGCTGCGATTAAAGCGGTAACCGCTGTTGCTTAACAGAACTGTTTTACCG 480

Qy 161 AlaSerProLeuTyr 165  
Db 481 CGCTCACCGTTGTATC 495

RESULT 3  
US-10-359-369-37  
; Sequence 37, Application US/10359369  
; Publication No. US20030215927A1  
; GENERAL INFORMATION:  
; APPLICANT: E. I. duPont de Nemours and Company, Inc.  
; APPLICANT: Viitanen, Paul  
; APPLICANT: Meyer, Knut  
; APPLICANT: Van Dyk, Drew  
; TITLE OF INVENTION: UDP-Glucosyltransferases  
; FILE REFERENCE: CUI821 US NA  
; CURRENT APPLICATION NUMBER: US/10/359.369  
; CURRENT FILING DATE: 2003-02-06  
; PRIOR APPLICATION NUMBER: 60/355,511  
; PRIOR FILING DATE: 2002-02-07  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 37  
; LENGTH: 495  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(495)  
US-10-359-369-37

Alignment Scores:  
Pred. No.: 1.91e-102 Length: 495  
Score: 854.00 Matches: 165  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 7 Gaps: 0

US-10-718-311-4 (1-165) x US-10-359-369-37 (1-495)

Qy	1	MetSerHisProAlaLeuThrGlnLeuAargAlaLeuAargTyrCysLysGluLeuProAla	20
Db	1	ATGTCAACCCCGCGTTAACGCAACTGGCTGGCTATGTGTAAGAGATCCCTGCC	60
Qy	21	LeuAspProGlnLeuLeuAspTrpLeuLeuGluLeuAaspSerMetThrLysAargPheGlu	40
Db	61	CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGA	120
Qy	41	GlnGlnGlyLysThrValSerValThrMetIleAargGluGlyPheValGluGlnAsnGlu	60
Db	121	CAGCAGGGAAAAACGGTAAGCTGACGATGATCCGCGAAGGGTTTGTCTGAGACGAAATGAA	180
Qy	61	IleProGluGluLeuProLeuLeuProLysGluSerAargTyrTrpLeuAargGluLeuLeu	80
Db	181	ATCCCCGAAGAACTGCCGCTGCTCCGAAGAAGACTCGTTACTGCTTACGTGAATTTTG	240
Qy	81	LeuCysAlaAspGlyGluProTrpLeuAlaGlyAargThrValValProValSerThrLeu	100
Db	241	TTATGTGCCGATGGTGAACCGTGGCTTGC CGCTCGTACCCGTCGTCTCTGTGTCAACGTTA	300
Qy	101	SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyAargTyrLeuPhe	120
Db	301	AGCGGCGCGAGCTGGCTTACAAAATTTGGTAAACCCGCTTAGGACCGCTATCTGTTTC	360
Qy	121	ThrSerSerThrLeuThrAargAspPheIleGluLeuGlyAargAspAlaGlyLeuTrpGly	140
Db	361	ACATCATCGACATTAAACCGGACCTTTATTGAGATAGCCCGTGATGCCGGGCTGTGGGG	420
Qy	141	ArgAqgSerArgLeuAargLeuSerClyLysProLeuLeuLeuThrGluLeuPheLeuPro	160
Db	421	CGACGTTCCCGCTCGATTAAAGCGGTAAACCGCTGTGTCTAAACAGAACTGTGTTTACC	480
Qy	161	AlaSerProLeuTyr	165
Db	481	GGGTACACCGTTGTAC	495

## RESULT 4

```

US-10-699-050-12
; Sequence 12: Application US/10699050
; Publication No. US2004014237A1
; GENERAL INFORMATION:
; APPLICANT: Flint, Dennis
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; TITLE OF INVENTION: Sinapoylglucose
; FILE REFERENCE: BC1034 US NA
; CURRENT APPLICATION NUMBER: US/10/6
; CURRENT FILING DATE: 2003-10-30
; PRIOR APPLICATION NUMBER: US/09/896
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/216,611
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 12
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-699-050-12

```

Alignment Scores:		
Pred. No.:	1,91e-102	Length:
Score:	854.00	Matches:
Percent Similarity:	100.0%	Conservative:
Best Local Similarity:	100.0%	Mismatches:
Query Match:	100.0%	Indels:
DB:	8	Gaps:

US-10-718-311-4 (1-165) x US-10-699-050-12 (1-495)

Qy	1	MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla	20
Db	1	ATGTCACACCCCGCGTTAAACCAACTGCGTGCCTGCCTATTGTATAACAGATCCCTGCC	60
Qy	21	LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu	40
Db	61	CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCCATGACAAAAAGTTTGA	120
Qy	41	GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAenGlu	60
Db	121	CAGCAGGGAAAAACGGTAGCGTGACATGATCCCGAAGGGTTGTGCGACGAATGAA	180
Qy	61	IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu	80
Db	181	ATCCCCGAAGAACTGCCGCTGCTGCCGAAGAGTCTCGTTACTGGTTACGTGAAATTTG	240
Qy	81	LeuCysAlaAspGlyGluProTrpTrpLeuAlaGlyArgThrValValProValSerThrLeu	100
Db	241	TTATGTGCCGATGGTGAACCGTGGCTTCGCCGTCGTACCGTCCTCTGTCAACGTTA	300
Qy	101	SerClyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe	120
Db	301	AGCGGGCCGAGCTGGCGGTTACAAAAATGGGTAAAAACGCCGTTAGACGCCTATCTGTC	360
Qy	121	ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly	140
Db	361	ACATCATCGACATTAAACCCGGGACCTTATTGAGATAGCCCGTGATGCCGGGCTGTGGGG	420
Qy	141	ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro	160
Db	421	CGAGCTTCCCGCCTCGGATTAAACGGGTAAACCGCTGTTGCTTAACAGAACTGTTTTACCG	480
Qy	161	AlaSerProLeuTyr	165
Db	481	CGCTCACCGTTGTAC	495

## RESULT 5

US-10-718-311-3

; Sequence 3, Application US/10718311

; Publication No. US20040143867A1

; GENERAL INFORMATION:

**; APPLICANT: VIITANEN, PAUL V.**

APPLICANT: MEYER, KNUT

APPLICANT: VAN DYK, DREW

; TITLE OF INVENTION: HIGH

; TITLE OF INVENTION: IN GREEN PLANTS

FILE REFERENCE: BC1015 US NA

; CURRENT APPLICATION NUMBER: US/10/718,311

; CURRENT FILING DATE: 2003-11-20

; PRIOR APPLICATION NUMBER: US/09/855,341

PRIOR FILING DATE: 2001-05-15

; NUMBER OF SEQ ID NOS: 16

SOFTWARE: MICROSOFT OFFICE 97

; SEQ ID NO 3  
1 ENCYCL 405

```

; LENGTH: 495
; TYPE: DNA

```

; TYPE: DNA  
; ORGANISM: Unknown Organism

ORGANISM	FEATURE
;	
:	

```
;; FEATURE: Description of Unknown Organism.E coli
```

; OTHER INFO  
IIS-10-718-311-3

Alignment Scores:

Pred. No.:	1,91e-102	Length:	495
Score:	854.00	Matches:	165
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	100.0%	Indels:	0
DB:	8	Gaps:	0

US-10-718-311-4 (1-165) x US-10-718-311-3 (1-495)

1 MetSerHisproAlaleuThrGlnLeuArqAlaleuArqTvrCvsLvsgluileProAla 20



```
|||||
Db 1 ATGTCACACCCCGGTTAAAGCAACTGCGTCGCGCTATTGTAAAGAGATCCCTGCC 60
Qy
21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db
61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGAA 120
Qy
41 GlnGlnGlyLysThrValSerValThrMetLeuArgGluGlyPheValGlnGlnGlu 60
Db
121 CAGCAGGGGAAAACCGTAAAGCGTGACGATATCCGCGAAGGTTTGTGAGCAGAGTAA 180
Qy
61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeu 80
Db
181 ATCCCGGAAGAACTGCGCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAATTTTG 240
Qy
81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db
241 TTATGTGCGGATGGTGAACCGTGGCTTCCCGTGGTACCGTCTGCTGTCAACGTTA 300
Qy
101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db
301 AGCGGGCGGAGCTGGCGCTTACAAAATTTGGTAAACCGCGTTAGGACGCTATCTGTT 360
Qy
121 ThrSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db
361 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCCCGTGATGCCGGGCTGTGGGG 420
Qy
141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db
421 CGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTACCG 480
Qy
161 AlaSerProLeuTyr 165
Db
481 CGGTCACCGTTGTAC 495
RESULT 6
US-10-462-162-29
; Sequence 29, Application US/10462162
; Publication No. US20040261147A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; APPLICANT: Flint, Dennis
; TITLE OF INVENTION: High Level Production of Arbutin in Green Plants and Microbes
; FILE REFERENCE: CL 2155 US NA
; CURRENT APPLICATION NUMBER: US/10/462,162
; CURRENT FILING DATE: 2003-06-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 29
; LENGTH: 498
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-462-162-29
Alignment Scores:
Pred. No.: 1.93e-102 Length: 498
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservativeness: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0
US-10-718-311-4 (1-165) x US-10-462-162-29 (1-498)
Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 1 ATGTCACACCCCGGTTAAAGCAACTGCGTCGCGCTATTGTAAAGAGATCCCTGCC 60
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGAA 120
Qy
21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db
61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGAA 120
```

```
Qy 41 GlnGlnGlyLysThrValSerValThrMetLeuArgGluGlyPheValGlnGlnGlu 60
Db 121 CAGCAGGGGAAAACCGTAAAGCGTGACGATATCCGCGAAGGTTTGTGAGCAGAGTAA 180
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeu 80
Db 181 ATCCCGGAAGAACTGCGCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAATTTTG 240
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 241 TTATGTGCGGATGGTGAACCGTGGCTTCCCGTGGTACCGTCTGCTGTCAACGTTA 300
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 301 AGCGGGCGGAGCTGGCGCTTACAAAATTTGGTAAACCGCGTTAGGACGCTATCTGTT 360
Qy 121 ThrSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
Db 361 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCCCGTGATGCCGGGCTGTGGGG 420
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 421 CGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTACCG 480
Qy 161 AlaSerProLeuTyr 165
Db 481 CGGTCACCGTTGTAC 495
RESULT 7
US-09-855-341-7
; Sequence 7, Application US/09855341
; Patent No. US20020002715A1
; GENERAL INFORMATION:
; APPLICANT: VIITANEN, PAUL V.
; APPLICANT: MEYER, KNUT
; APPLICANT: VAN DYK, DREW
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID
; FILE REFERENCE: IN GREEN PLANTS
; FILE REFERENCE: BC1015 US NA
; CURRENT APPLICATION NUMBER: US/09/855,341
; CURRENT FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 7
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:synthetic CPL
US-09-855-341-7
Alignment Scores:
Pred. No.: 3e-102 Length: 684
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservativeness: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0
US-10-718-311-4 (1-165) x US-09-855-341-7 (1-684)
Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
Db 187 ATGTCACACCCCGGTTAAAGCAACTGCGTCGCGCTATTGTAAAGAGATCCCTGCC 246
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db 247 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGAA 306
Qy 41 GlnGlnGlyLysThrValSerValThrMetLeuArgGluGlyPheValGlnGlnGlu 60
Db 307 CAGCAGGGGAAAACCGTAAAGCGTGACGATATCCGCGAAGGTTTGTGAGCAGAGTAA 366
```



```

QY 61 IleProGluGluLeuProLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
|
|
|
Db 367 ATCCCCGAAGAAGACTGCGCTGCGCGAAGAGTCTCGTACTGGTTACGTGAATTTTG 426
|
|
|
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
|
|
|
Db 427 TTATGTGCGCATGGTGAACCGTGGCTTGGCGGTGCTACCGTCTGCTGTCAACGTTA 486
|
|
|
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
|
|
|
Db 487 AGCGGCGCGAGCTGGCGTTACAAAATTGGGTAAACCGCGTTAGACGCTATCTGTTTC 546
|
|
|
QY 121 ThrSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
|
|
|
Db 547 ACATCATCGACATTAAACCGGACCTTTATTGAGTAGGCGCGTATCGCGGCTGTGGGG 606
|
|
|
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 160
|
|
|
Db 607 CGACGTTCCCGCTCGCATTAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTACCG 666
|
|
|
QY 161 AlaSerProLeuTyr 165
|
|
|
Db 667 GCGTCACCGTTGTAC 681

```

## RESULT 8

US-09-896-866B-16

; Sequence 16, Application US/09896866B

; Patent No. US20020151002A1

; GENERAL INFORMATION:

; APPLICANT: Flint, Dennis

; APPLICANT: Meyer, Knut

; APPLICANT: Viitanen, Paul

; TITLE OF INVENTION: Sinapoylglucose-Malate Sinapoyltransferase Form Malate Conjugates

; FILE OF INVENTION: Benzoic Acid Glucosides

; FILE REFERENCE: BC1034 US NA

; CURRENT APPLICATION NUMBER: US/09/896,866B

; CURRENT FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: 60/216,615

; PRIOR FILING DATE: 2000-07-07

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 16

; LENGTH: 684

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; NAME/KEY: misc feature

; OTHER INFORMATION: open reading frame of the chloroplast-targeted CPL fusion protein

US-09-896-866B-16

Alignment Scores:

Pred. No.:	3e-102	Length:	684
Score:	854.00	Matches:	165
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	100.0%	Indels:	0
DB:	3	Gaps:	0

US-10-718-311-4 (1-165) x US-09-896-866B-16 (1-684)

```

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
|
|
|
Db 187 ATGTGCACACCGCGTTAACGCAACTGCGTGGCTGCGCTATTGTAAAGAGATCCCTGCC 246
|
|
|
QY 21 LeuAppProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
|
|
|
Db 247 CTGGATCCCAACTGCTCGACTGGCTGTTGCTGGAGATTCCATGACAAAACGTTTGA 306
|
|
|
QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAnGlu 60
|
|
|
Db 307 CAGCAGGGGAAAAACGGTAAGCGTAGCATGATCCCGAAGGGTGTTCGAGCAGAAATGAA 366
|
|
|
QY 61 IleProGluGluLeuProLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80

```

```

QY 61 IleProGluGluLeuProLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
|
|
|
Db 367 ATCCCCGAAGAAGACTGCGCTGCGCGAAGAGTCTCGTACTGGTTACGTGAATTTTG 426
|
|
|
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
|
|
|
Db 427 TTATGTGCGCATGGTGAACCGTGGCTTGGCGGTGCTACCGTCTGCTGTCAACGTTA 486
|
|
|
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
|
|
|
Db 487 AGCGGCGCGAGCTGGCGTTACAAAATTGGGTAAACCGCGTTAGACGCTATCTGTTTC 546
|
|
|
QY 121 ThrSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140
|
|
|
Db 547 ACATCATCGACATTAAACCGGACCTTTATTGAGTAGGCGCGTATCGCGGCTGTGGGG 606
|
|
|
QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 160
|
|
|
Db 607 CGACGTTCCCGCTCGCATTAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTACCG 666
|
|
|
QY 161 AlaSerProLeuTyr 165
|
|
|
Db 667 GCGTCACCGTTGTAC 681

```

## RESULT 9

US-10-359-369-41

; Sequence 41, Application US/10359369

; Publication No. US20030215927A1

; GENERAL INFORMATION:

; APPLICANT: E. I. duPont de Nemours and Company, Inc.

; APPLICANT: Viitanen, Paul

; APPLICANT: Meyer, Knut

; APPLICANT: Van Dyk, Drew

; TITLE OF INVENTION: UDP-Glucosyltransferases

; FILE REFERENCE: C11821 US NA

; CURRENT APPLICATION NUMBER: US/10/359,369

; CURRENT FILING DATE: 2003-02-06

; PRIOR APPLICATION NUMBER: 60/355,511

; PRIOR FILING DATE: 2002-02-07

; NUMBER OF SEQ ID NOS: 46

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 41

; LENGTH: 684

; TYPE: DNA

; ORGANISM: Escherichia coli

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(684)

US-10-359-369-41

Alignment Scores:

Pred. No.:	3e-102	Length:	684
Score:	854.00	Matches:	165
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	100.0%	Indels:	0
DB:	7	Gaps:	0

US-10-718-311-4 (1-165) x US-10-359-369-41 (1-684)

```

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
|
|
|
Db 187 ATGTGCACACCGCGTTAACGCAACTGCGTGGCTGCGCTATTGTAAAGAGATCCCTGCC 246
|
|
|
QY 21 LeuAppProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40
|
|
|
Db 247 CTGGATCCCAACTGCTCGACTGGCTGTTGCTGGAGATTCCATGACAAAACGTTTGA 306
|
|
|
QY 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAnGlu 60
|
|
|
Db 307 CAGCAGGGGAAAAACGGTAAGCGTAGCATGATCCCGAAGGGTGTTCGAGCAGAAATGAA 366
|
|
|
QY 61 IleProGluGluLeuProLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80

```

Db 367 ATCCCGAGAACTGCGCTGCTGCCGAAAGAGTCTGTTACTGTTACGTGAATTTTG 426  
 Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100  
 Db 427 TTATGTGCGGATGGTGAACCGTGGCTTCCCGGTGCTACCGTCTGTTCTGTCAACGTTA 486  
 Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120  
 Db 487 AGCGGGCGGAGCTGGCGTTACAAAATTGGGTAAACCGCGTTAGGACGCTATCTGTTC 546  
 Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluLeuGlyArgAspAlaGlyLeuTrpGly 140  
 Db 547 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGCCCGTGTATGCCGGCTGTGGGG 606  
 Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160  
 Db 607 CGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTGCTTAACAGAACTGTTTTACCG 666  
 Qy 161 AlaSerProLeuTyr 165  
 Db 667 GCGTCACCGTTGTAC 681

## RESULT 10

US-10-699-050-16  
 ; Sequence 16, Application US/10699050  
 ; Publication No. US20040142437A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Flint, Dennis  
 ; APPLICANT: Meyer, Knut  
 ; APPLICANT: Viitanen, Paul  
 ; TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugates  
 ; TITLE OF INVENTION: Benzoic Acid Glucosides  
 ; FILE REFERENCE: BC1034 US NA  
 ; CURRENT APPLICATION NUMBER: US/10/699,050  
 ; CURRENT FILING DATE: 2003-10-30  
 ; PRIOR APPLICATION NUMBER: US/09/896,866B  
 ; PRIOR FILING DATE: 2001-06-29  
 ; PRIOR APPLICATION NUMBER: 60/216,615  
 ; PRIOR FILING DATE: 2000-07-07  
 ; NUMBER OF SEQ ID NOS: 19  
 ; SOFTWARE: Microsoft Office 97  
 ; SEQ ID NO 16  
 ; LENGTH: 684  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; OTHER INFORMATION: open reading frame of the chloroplast-targeted CPL fusion protein  
 US-10-699-050-16

## Alignment Scores:

Pred. No.: 3e-102 Length: 684  
 Score: 854.00 Matches: 165  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 100.0% Indels: 0  
 DB: 8 Gaps: 0

US-10-718-311-4 (1-165) x US-10-699-050-16 (1-684)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20  
 Db 187 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGCTATTGTTAAAGAGATCCCTGCC 246  
 Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40  
 Db 247 CTGGATCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGA 306  
 Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60  
 Db 307 CAGCAGGAAAAACCGGTAAAGCGTGACGATGATCCGCGAAGGGTTTGTGCGAGCAGATGAA 366

Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80  
 Db 367 ATCCCGAGAACTGCGCTGCTGCCGAAAGAGTCTGTTACTGTTACGTGAATTTTG 426  
 Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100  
 Db 427 TTATGTGCGGATGGTGAACCGTGGCTTCCCGGTGCTACCGTCTGTTCTGTCAACGTTA 486  
 Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120  
 Db 487 AGCGGGCGGAGCTGGCGTTACAAAATTGGGTAAACCGCGTTAGGACGCTATCTGTTC 546  
 Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluLeuGlyArgAspAlaGlyLeuTrpGly 140  
 Db 547 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGCCCGTGTATGCCGGCTGTGGGG 606  
 Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160  
 Db 607 CGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTGCTTAACAGAACTGTTTTACCG 666  
 Qy 161 AlaSerProLeuTyr 165  
 Db 667 GCGTCACCGTTGTAC 681

## RESULT 11

US-10-718-311-7  
 ; Sequence 7, Application US/10718311  
 ; Publication No. US20040143867A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Viitanen, Paul V.  
 ; APPLICANT: MEYER, KNUT  
 ; APPLICANT: VAN DYK, DREW  
 ; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZZOIC ACID  
 ; TITLE OF INVENTION: IN GREEN PLANTS  
 ; FILE REFERENCE: BC1015 US NA  
 ; CURRENT APPLICATION NUMBER: US/10/718,311  
 ; CURRENT FILING DATE: 2003-11-20  
 ; PRIOR APPLICATION NUMBER: US/09/855,341  
 ; PRIOR FILING DATE: 2001-05-15  
 ; NUMBER OF SEQ ID NOS: 16  
 ; SOFTWARE: MICROSOFT OFFICE 97  
 ; SEQ ID NO 7  
 ; LENGTH: 684  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence:synthetic CPL  
 US-10-718-311-7

## Alignment Scores:

Pred. No.: 3e-102 Length: 684  
 Score: 854.00 Matches: 165  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 100.0% Indels: 0  
 DB: 8 Gaps: 0

US-10-718-311-4 (1-165) x US-10-718-311-7 (1-684)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20  
 Db 187 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGCTATTGTTAAAGAGATCCCTGCC 246  
 Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40  
 Db 247 CTGGATCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGA 306  
 Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60  
 Db 307 CAGCAGGAAAAACCGGTAAAGCGTGACGATGATCCGCGAAGGGTTTGTGCGAGCAGATGAA 366  
 Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80

Db 367 ATCCCGAAGAACTCGCGTGTGCGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTTG 426  
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100  
Db 427 TTATGTGCCGATGGTGAACCGTGGCTTGGCGGTCTGACCGTCTGCTGTCTGTCACGTTA 486  
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120  
Db 487 AGCGGCGCGAGCTGCGGTTACAAAATTTGGGTAAACCGCGTTAGACGCTATCTGTTC 546  
Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140  
Db 547 ACATCATCGACATTAAACCGGACTTTATTGAGTAGGCGGTATCCCGGCTGTGGGG 606  
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 160  
Db 607 CGACGTTCCCGCTGGATTAAACCGGTAAACCGCTGTGCTAACAGAACTGTTTTACCG 666  
Qy 161 AlaSerProLeuTyr 165  
Db 667 GCGTCACCGTTGTAC 681

## RESULT 12

US-10-462-162-33  
; Sequence 33, Application US/10462162  
; Publication NO. US20040261147A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Meyer, Knut  
; APPLICANT: Viitanen, Paul  
; APPLICANT: Flint, Dennis  
; TITLE OF INVENTION: High Level Production of Arbutin in Green Plants and Microbes  
; FILE REFERENCE: CL 2155 US NA  
; CURRENT APPLICATION NUMBER: US/10/462,162  
; CURRENT FILING DATE: 2003-06-16  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 33  
; LENGTH: 684  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
US-10-462-162-33

Alignment Scores:  
Pred. No.: 3e-102 Length: 684  
Score: 854.00 Matches: 165  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 9 Gaps: 0

US-10-718-311-4 (1-165) x US-10-462-162-33 (1-684)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 20  
Db 187 ATGTACACCCCGCGTTAACGCAACTGCGTGGCTATTTAAAGAGATCCCTGCC 246  
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40  
Db 247 CTGGATCCGCAACTGCTGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTTGAA 306  
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAenGlu 60  
Db 307 CAGCAGGGGAAAAACGGTAAGCTGACGATGATCCCGAAGGTTGTGCGACGAGATGAA 366  
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeu 80  
Db 367 ATCCCGGAAGAACTCGCGTGTGCGGTAAGAGTCTCGTTACTGTTACGTGAAATTTTG 426  
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100  
Db 427 TTATGTGCCGATGGTGAACCGGTGCGGTGCTGACCGTCTGCTGTGTCACGTTA 486

Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120  
Db 487 AGCGGCGCGAGCTGCGGTTACAAAATTTGGGTAAACCGCGTTAGACGCTATCTGTTC 546  
Qy 121 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 140  
Db 547 ACATCATCGACATTAAACCGGACTTTATTGAGTAGGCGGTATCCCGGCTGTGGGG 606  
Qy 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 160  
Db 607 CGACGTTCCCGCTGGATTAAACCGGTAAACCGCTGTGCTAACAGAACTGTTTTACCG 666  
Qy 161 AlaSerProLeuTyr 165  
Db 667 GCGTCACCGTTGTAC 681

## RESULT 13

US-10-462-162-47  
; Sequence 47, Application US/10462162  
; Publication NO. US20040261147A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Meyer, Knut  
; APPLICANT: Viitanen, Paul  
; APPLICANT: Flint, Dennis  
; TITLE OF INVENTION: High Level Production of Arbutin in Green Plants and Microbes  
; FILE REFERENCE: CL 2155 US NA  
; CURRENT APPLICATION NUMBER: US/10/462,162  
; CURRENT FILING DATE: 2003-06-16  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 47  
; LENGTH: 3452  
; TYPE: DNA  
; ORGANISM: artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Nucleotide sequence of nucleic acid fragment encoding CPL (from  
; OTHER INFORMATION: Escherichia coli), pHBA 1-hydroxylase (from Candida  
; OTHER INFORMATION: parapsilosis), and UGT72B1 (from Arabidopsis thaliana).  
US-10-462-162-47

Alignment Scores:  
Pred. No.: 2.83e-101 Length: 3452  
Score: 854.00 Matches: 165  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 9 Gaps: 0

US-10-718-311-4 (1-165) x US-10-462-162-47 (1-3452)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 20  
Db 4 ATGTACACCCCGCGTTAACGCAACTGCGTGGCTATTTAAAGAGATCCCTGCC 63  
Qy 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40  
Db 64 CTGGATCCGCAACTGCTGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTTGAA 123  
Qy 41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAenGlu 60  
Db 124 CAGCAGGGGAAAAACGGTAAGCTGACGATGATCCCGAAGGTTGTGCGACGAGATGAA 183  
Qy 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeu 80  
Db 184 ATCCCGGAAGAACTCGCGTGTGCGGTAAGAGTCTCGTTACTGTTACGTGAAATTTTG 243  
Qy 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100  
Db 244 TTATGTGCCGATGGTGAACCGGTGCGGTGCTGACCGTCTGCTGTGTCACGTTA 303  
Qy 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120



```

Db      952 TAAGCGGCGGAGCTGGCGTTACAAAAATTGGGTAAAAACGCCGTTAGGACGCTATCTGT 1011
QY      120 heThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpG 140
        |||||
Db      1012 TCACATCATCGACATTAAACCCGGGACTTTATTGAGATAGGCCCGTGATGCCGGGCTGTGGG 1071
        |||||
QY      140 lyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuP 160
        |||||
Db      1072 GCGGACGTTCCCGCCTGCGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTAC 1131
        |||||
QY      160 rcAlaSerProLeuTyr 165
        |||||
Db      1132 CGGCGTCACCGGTGTAC 1148
        |||||

```

Search completed: May 30, 2006, 03:42:00  
 Job time : 1139.79 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 30, 2006, 03:03:47 ; Search time 34.9701 Seconds  
(without alignments)  
798.935 Million cell updates/sec

Title: US-10-718-311-4

Perfect score: 854

Sequence: 1 MSHPALQALRYLCKEIPA.....RLSGKPLLLTELFLPASPLY 165

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 242596 seqs, 56442199 residues

Total number of hits satisfying chosen parameters: 485192

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlh  
-Q=/abss/ABSSWEB\_spool/US10718311/runat\_26052006\_164936\_13807/app\_query.fasta\_1  
-DB=Published Applications NA New -QFMT=fastap -SUFFIX=rnpbn -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-MAXLEN=2000000000 -HOST=absg03h  
-USER=US10718311 @CGN1\_1\_26 @runat\_26052006\_164936\_13807 -NCPU=6 -ICPU=3  
-NO\_WMAP -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.New.\*

- 1: /ENC\_Celerra\_SIDS3/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 2: /ENC\_Celerra\_SIDS3/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 3: /ENC\_Celerra\_SIDS3/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 4: /ENC\_Celerra\_SIDS3/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 5: /ENC\_Celerra\_SIDS3/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 6: /ENC\_Celerra\_SIDS3/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 7: /ENC\_Celerra\_SIDS3/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*
- 8: /ENC\_Celerra\_SIDS3/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80	9.4	2307	US-11-293-697-1067	Sequence 1067, Ap
2	76.5	9.0	3350	US-11-293-697-509	Sequence 509, App
3	76	8.9	1104	US-11-217-529-2348	Sequence 2348, Ap
4	75.5	8.8	601	US-10-488-619-1930	Sequence 1930, Ap
5	75.5	8.8	1791	US-11-217-529-1938	Sequence 1938, Ap
6	74.5	8.7	2384	US-11-293-697-1066	Sequence 1066, Ap
7	73	8.5	489	US-11-217-529-174332	Sequence 174332, Ap
8	73	8.5	1406	US-10-953-349-34874	Sequence 34874, A
9	72	8.4	3023	US-11-293-697-665	Sequence 665, App

10	71.5	8.4	966	6	US-10-524-433-1	Sequence 1, Appli
11	71.5	8.4	2490	7	US-11-293-697-200	Sequence 200, App
12	71.5	8.4	4534	7	US-10-473-6918-1	Sequence 1, Appli
13	71	8.3	3105	7	US-11-293-697-336	Sequence 336, App
14	70	8.2	1191	7	US-11-217-529-77335	Sequence 77335, A
15	70	8.2	2555	6	US-10-523-014-2	Sequence 2, Appli
16	70	8.2	3516	7	US-11-217-529-1696	Sequence 1696, Ap
17	69.5	8.1	1641	7	US-11-217-529-76487	Sequence 76487, A
18	69.5	8.1	138941	6	US-10-489-730-10	GENERAL INFORMATI
19	69	8.1	2019	7	US-11-293-697-1629	Sequence 1629, Ap
20	68.5	8.0	1887	7	US-11-217-529-3713	Sequence 3713, Ap
21	68.5	8.0	3315	7	US-11-217-529-76752	Sequence 76752, A
22	68	8.0	1104	7	US-11-217-529-78541	Sequence 78541, A
23	68	8.0	1365	6	US-10-953-349-21628	Sequence 21628, A
24	68	8.0	1484	6	US-10-953-349-20563	Sequence 20563, A
25	68	8.0	2798	6	US-10-953-349-40225	Sequence 40225, A
26	67.5	7.9	851	6	US-10-953-349-37006	Sequence 37006, A
27	67.5	7.9	1314	6	US-10-953-349-22504	Sequence 22504, A
28	67.5	7.9	13128	7	US-11-118-524-1	Sequence 1, Appli
29	67.5	7.9	128361	6	US-10-505-928-151	Sequence 151, App
30	67	7.8	987	7	US-11-226-605-68	Sequence 68, Appli
31	67	7.8	1422	7	US-11-217-529-75461	Sequence 75461, A
32	67	7.8	3075	7	US-11-226-605-51	Sequence 51, Appli
33	67	7.8	3157	6	US-10-468-193-29	Sequence 29, Appli
34	67	7.8	3252	7	US-11-217-529-76569	Sequence 76569, A
35	67	7.8	10905	6	US-10-505-928-568	Sequence 568, App
36	66.5	7.8	1638	7	US-11-217-529-876	Sequence 876, App
37	66	7.7	614	7	US-11-318-813-8	Sequence 8, Appli
38	66	7.7	650	6	US-10-953-349-7808	Sequence 7808, Ap
39	66	7.7	895	6	US-10-953-349-4803	Sequence 4803, Ap
40	66	7.7	1824	6	US-10-953-349-1041	Sequence 1041, Ap
41	66	7.7	2030	7	US-11-322-999-5	Sequence 5, Appli
42	66	7.7	2424	7	US-11-217-529-75664	Sequence 75664, A
43	66	7.7	8764	7	US-11-322-999-137	Sequence 137, App
44	66	7.7	26850	7	US-11-318-813-41	Sequence 41, Appli
45	65.5	7.7	613	6	US-10-488-619-2190	Sequence 2190, Ap

ALIGNMENTS

RESULT 1  
US-11-293-697-1067  
; Sequence 1067, Application US/11293697  
; Publication No. US20060105376A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/11/293,697  
; CURRENT FILING DATE: 2005-12-05  
; PRIOR APPLICATION NUMBER: US/10/108,260  
; PRIOR FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1067  
; LENGTH: 2907  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-293-697-1067

Alignment Scores:  
Pred. No.: 5.18 Length: 2907  
Score: 80.00 Matches: 42  
Percent Similarity: 42.9% Conservative: 18  
Best Local Similarity: 30.0% Indels: 55  
Query Match: 9.4% Gaps: 8  
DB:

US-10-718-311-4 (1-165) x US-11-293-697-1067 (1-2907)

Qy 32 GluapSerMetThrLysArgPheGluGlnGlnGlySerValThrMetIle 51  
DB 559 GATGATAAATTGTCAAAAGCCACCAAGAGTCAGGACAGGAACTCAGTTTCTCCCCAGCTG 618

```
Qy 52 Arg-----GluGlyPhe-----ValGluGlnAenGluLeuProGluGluLeuProLeu 67
Db 619 AGGGCTGTGCAGGGATTTCAAGGTACCTGTGCGAAACCCAGCGCTCACTATTGAACCTCA 678
Qy 68 LeuProLysGluSerArgTyrTrpLeuArgGluLeuLeuLeuCysAlaAspGlyGluPro 87
Db 679 GTCAGAAACCAACCCCGTCAGTGGATTTAGAAA-----TGTGACAGGGAT--- 723
Qy 88 TrpLeuAlaGlyArgThrValProValSerThrLeuSerGlyProGluLeuAlaLeu 107
Db 724 ---TTCAGAGGGCAATGGTCCAGCCC-----TCACAACCCAGCACCACTCTG 768
Qy 108 GlnLysLeuGlyLysThrProLeuGly-Arg-----TyrLeuPheThrSerThrLe 125
Db 769 CAGGCATGGAACAGCAGGCCCCAGAGGACAGGGGCTTCTCAATATGGGCACACAGCACCA 828
Qy 125 uThrArgAspPheIleGluLeu-----GlyArgAspAlaGlyLeuTr 139
Db 829 GACACTCCATAATAGAAAGTGAATATCCAAAGTCCCACTCCAGGAAGCTGACCACTG 888
Qy 139 p-----GlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGlu 156
Db 889 GGTGATGACTCCCCAGCGTCCAAACCCCTCTCTGCGAGTCCCTTGCTGCTCCAGGAG 946

RESULT 2
US-11-293-697-509/c
; Sequence 509, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; PRIOR FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 509
; LENGTH: 3550
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-509

Alignment Scores:
Pred. No.: 18.3 Length: 3550
Score: 76.50 Matches: 43
Percent Similarity: 41.1% Conservative: 15
Best Local Similarity: 30.5% Mismatches: 35
Query Match: 9.0% Indels: 49
DB: 7 Gaps: 8

US-10-718-311-4 (1-165) x US-11-293-697-509 (1-3550)

Qy 27 AspTrpLeuLeuGluAspSerMetThrLysArgPheGluGlnGlnGlyLysThrVal 46
Db 397 GACAACACCCCTAAGTGAAGAGTGCCTCACAACAGAGTCCACAGAGCGGCAAAAT--GCG 340
Qy 47 SerValThrMetIleArgGluGlyPheValGluGln-----Asn 59
Db 339 AGTGTG-----GAGGAAGCATCTCCAGGATTCAGGGCGCCCGCTGACGAC 292
Qy 60 GluLeuProGluGluLeuProLeuProLysGluSerArgTyrTrpLeuArgGluLeu 79
Db 291 TCCTGCCGGCAAAAGTCCCCTGCTCCAGCAGAATGAGACGATGCCCTGACCCCATC 232
Qy 80 -----LeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValPro---- 96
Db 231 GTCAGGTGACCTGC-----TGG-----CCAACCTCAGTTTCTCCAAGG 193
Qy 97 ValSerThrLeuSerGlyProGluLeuAla----- 106
```

```
Db 192 GTCACAATGCTCTTTGGTCTCCTGAGTTACCTCATCGGGAGACGAGAGGCATGGGAGGAAA 133
Qy 107 -----LeuGlnLysLeuGlyLysThrProLeu 115
Db 132 AGGGGGTACAGACCCGAGGAGGAGTTGGCTTCAGACAGAAAAGACAAGAAACCCGGGTG 73
Qy 116 GlyArgTyrLeuPhe-----ThrSerSerThrLeuThrArgAspPhe 129
Db 72 GGAAGATGGGCTTTCTGGGGGTAGGAATCACACACTTCGTCACACTGTGACAAAAGACAGA 13
Qy 130 Ile 130
Db 12 ATT 10

RESULT 3
US-11-217-529-2348
; Sequence 2348, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 2348
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-2348

Alignment Scores:
Pred. No.: 3.89 Length: 1104
Score: 76.00 Matches: 42
Percent Similarity: 38.3% Conservative: 28
Best Local Similarity: 23.0% Mismatches: 63
Query Match: 8.9% Indels: 50
DB: 7 Gaps: 10

US-10-718-311-4 (1-165) x US-11-217-529-2348 (1-1104)

Qy 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeu---Pro 19
Db 547 GTAGATAGACCCATAATTTCTTCAATCACCAGGAGTAGCGTACTCTACAAGAGTAGTCCG 606
Qy 20 AlaLeuAspProGlnLeuLeuAspTrp-----LeuLeuLeuGlu 32
Db 607 TCACCTCGTAATAGATGATCGCGTCCCAAAAAGATTGTAAGAGGACCTCAATGTGAG 666
Qy 33 AspSerMetThrLys-----ArgPheGluGlnGlnGlyLysThrVal-----Ser 47
Db 667 GTTAACAATAACAGCAGATGTTTGGAGAGCGGCAAAATTCAGGTAAAGAGTCCAGGCTGG 726
Qy 48 ValThrMetIleArgGluGly-----PheValGluGln----- 58
Db 727 ATAACCTTTGGTGGCTGAATCCACAGCAAAAAGTTGGAGCTATTTCCGCCGAAGATATTGG 786
Qy 59 -----AsnGluLeuProGluLeuProLeuProLysGluSerArgTyrTrpLeu 76
Db 787 GATCAGCGCTCTATACCCGAGAACTTGGT-----GAAAAAGTTGCTTACCGTTTATTA 840
Qy 77 ArgGluLeuLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValPro 96
Db 841 GAAGAAATATCGAAAAGCGCT-----GCTGTCGGTAGAAGTCAACTTCCA 885
```



```
QY 97 ValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGly 116
Db 886 CTAGCA-----ATCGTATACATGCTCGAAGAAAGAGATATCGGC 927
QY 117 ArgTyrLeuPheThrSerThrLeuThrArgAspPheIleGluLeuGlyArgAspAla 136
Db 928 AGATTAAAGATTAAAGACCCAGATAGATGAAGAGTTTATAACCCCTCTTGAGAGATATA 987
QY 137 GlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGlu 156
Db 988 -----AAGAGATCTTCTGCACAGAA 1008
QY 157 LeuPheLeu 159
Db 1009 GTCTTTTG 1017
RESULT 4
US-10-488-619-1930
; Sequence 1930, Application US/10488619
; Publication No. US20060099578A1
; GENERAL INFORMATION:
; APPLICANT: Greenlee, Winner and Sullivan, P. C.
; TITLE OF INVENTION: Human Mitochondrial DNA Polymorphisms, Haplogroups, Associations
; FILE REFERENCE: 98-01 WO
; CURRENT APPLICATION NUMBER: US/10/488,619
; PRIORITY FILING DATE: 2004-03-01
; NUMBER OF SEQ ID NOS: 3040
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1930
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-488-619-1930
Alignment Scores:
Pred. No.: 1.85 Length: 601
Score: 75.50 Matches: 32
Percent Similarity: 47.2% Conservative: 28
Best Local Similarity: 25.2% Mismatches: 38
Query Match: 8.8% Indels: 29
DB: 6 Gaps: 9
US-10-718-311-4 (1-165) x US-10-488-619-1930 (1-601)
QY 10 ArgAlaLeuArg-----TyrCysLysGluIleProAlaLeuAsp 22
Db 151 CGTTCCTTAAGAGAAAGCCATCAGAGAAGTGTATTGCCACCATCAAGTCCACCACT 210
QY 23 ProGlnLeuLeuAsp-----TrpLeuLeuLeuGluAspSerMetThrLys 37
Db 211 GCTGAGCTCTACCATACAGACCCATGGCTGGGCTCCAGCTTTGATGATCTTTAAG 270
QY 38 ArgPheGluGlnGlnGlyThrValSer-----ValThrMetIleArgGluGly 54
Db 271 AATTTTAAAGGAAAGCAAAATCATCTCTGAATATGATCATCACCTTGATGACT--- 327
QY 55 PheValGluGlnAsnGluIleProGluGluLeuProLeuProLysGluSerArgTyr 74
Db 328 TACATAGAGGAAATAAAGCTCAGAAAGCTGTTCTGTAATTTGAAGAGTA----- 378
QY 75 TrpLeuArgGluIle-----LeuLeuCysAlaAspGlyGluProTrpLeu 89
Db 379 ---CTGAGAGACATCGAGAGTCTCTCTGCACATAGCTGTGACAGGGGAAACAGGC--- 432
QY 90 AlaGlyArgThrValProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLys 109
Db 433 GCAGGGAAGTCCACTTTC---ATCAATACCTTGAGGGG-----GTGGGGCATGAAGAA 483
QY 110 LeuGlyLysThrProLeuGly 116
Db 484 AAAGGTGCAGCCCCCCTACTGCG 504
```

```
RESULT 5
US-11-217-529-1938
; Sequence 1938, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; PRIORITY FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1938
; LENGTH: 1791
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-1938
Alignment Scores:
Pred. No.: 8.99 Length: 1791
Score: 75.50 Matches: 36
Percent Similarity: 34.3% Conservative: 10
Best Local Similarity: 26.9% Mismatches: 41
Query Match: 8.8% Indels: 47
DB: 7 Gaps: 8
US-10-718-311-4 (1-165) x US-11-217-529-1938 (1-1791)
QY 62 ProGluGluLeuPro-----LeuLeuPro 69
Db 94 CCAGATGACTTCCCGACGTGGTGAAGCTGTTAAGGAACACCGTTGGCAGCTGGAGCCC 153
QY 70 LysGluSerArgTyrTrpLeuArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeu 89
Db 154 GCATTGAGCAGGTATTTC-----GACGGGAAA---TGGAGG 186
QY 90 AlaGly-----ArgThrValValProValSerThrLeu 100
Db 187 GATGCTCCGAGGAGGGATGGAGAGCCCACTGCCACAGCCAGCCCAATCCACGCC 246
QY 101 SerGlyProGluLeuAlaLeuGlnLys-----LeuGlyLysThrProLeuGlyArgTyr 118
Db 247 ATGGCAGAGGAGTTCGCTGGGACTCCGCCGCTGGTCCCGAGGCTCTATCA----- 300
QY 119 LeuPheThrSer-----SerThrLeuThrArgAspPhe 129
Db 301 ---TTTACAGCTGCATTCGCCGTAGTCAGACCAATTCGCCAGCGAGCTTCGCTCATGATTTT 357
QY 130 IleGluIle-----GlyArgAspAlaGlyLeuTrpGlyArgSerArgLeuArg 146
Db 358 AGAACCATTTGGCTGAACGACGCTCCCAATACCGTATGTCATGTCATGTCCTTCAGT 417
QY 147 LeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db 418 TACAACGGTAACCCCTTTCTTTTGTGTTGCTGCTGATCCCG 459
RESULT 6
US-11-293-697-1066
; Sequence 1066, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; PRIORITY FILING DATE: 2005-12-05
```

```
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1066
; LENGTH: 2384
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-1066

Alignment Scores:
Pred. No.: 17.9 Length: 2384
Score: 74.50 Matches: 25
Percent Similarity: 48.3% Conservative: 4
Best Local Similarity: 41.7% Mismatches: 21
Query Match: 8.7% Indels: 10
DB: 7 Gaps: 2

US-10-718-311-4 (1-165) x US-11-293-697-1066 (1-2384)
Qy 65 LeuProLeu-LeuProLysGluSerArgTyrTrpLeuArgGluLeuLeuCysAlaAs 84
Db 11 TTGCCACTCCCTTGGCAAGGAGCGGAGACTCAGTTCCCGGGGCTCTTGGCGGCAC 70
Qy 84 pGlyGluProTrpLeuAlaGlyArgThrValValPro---ValSerThrLeuSerGlyPr 103
Db 71 AGGTGAGCCCTGGCTGCGCGCGCCCTCTCTCCCGCGGCTCCAGATGGGGGTCC 130
Qy 103 oGluLeuAla-----LeuGlnLysLeuGlyLysThrPro 114
Db 131 GGAGTGGCGCCAGGCTCTGAGTACCTAGTCTGCAGACTAGCGGGCATTTGGCCA 188

RESULT 7
US-11-217-529-174332
; Sequence 174332, Application US/11217529
; Publication No. US2006099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 174332
; LENGTH: 489
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-174332

Alignment Scores:
Pred. No.: 2.76 Length: 489
Score: 73.00 Matches: 39
Percent Similarity: 34.3% Conservative: 7
Best Local Similarity: 29.1% Mismatches: 56
Query Match: 8.5% Indels: 32
DB: 7 Gaps: 6

US-10-718-311-4 (1-165) x US-11-217-529-174332 (1-489)
Qy 51 IleArgGluGlyPheValGluGlnAsnGluLeuProGluGluLeuPro----- 66
Db 31 TTAATGATGATCCAGGTATCATCAGCAATTTCCCGCCAGAGAGACTTGCAGATGTGTGTAGA 90
Qy 67 -----LeuLeuProLysGluSerArgTyrTrpLeuArgGlu 78
```

```
Db 91 CTGTTGAGGAACCATGGTTGGCACTTAGAACCAACGATTGACCGGATATTTTCGATGGAGAG 150
Qy 79 IleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSer 98
Db 151 -----TGGAAAGGCGAACCAGACAGATGGGAGAGTCC---ACTCAAACGTCT 195
Qy 99 ThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyr 118
Db 196 ACACCCATGGCTGAGACCTTGGTTCTCTCCCGCTTGGGACCTAGACCCGTTG----- 246
Qy 119 LeuPheThrSerSer-----ThrLeuThrArgAspPhe 129
Db 247 TTAATTACAGCTCAGTCCCGTAGTCAGACCATTCAGGCAACTTCGGAATGATTTC 306
Qy 130 IleGluLeu-----GlyArgAspAlaGlyLeuTrpGlyArgSerArgLeuArg 146
Db 307 AGAACAAATGGTTAAATGGAGCGCTCTAACTGTATGTCATGTTTGAATCTTTTAGT 366
Qy 147 LeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 160
Db 367 TATGATGGTAACCCCTTTCTTTTATTACTGTGATCCCT 408

RESULT 8
US-10-953-349-34874/C
; Sequence 34874, Application US/10953349
; Publication No. US2006010745A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 34874
; LENGTH: 1406
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
US-10-953-349-34874

Alignment Scores:
Pred. No.: 12.7 Length: 1406
Score: 73.00 Matches: 29
Percent Similarity: 42.0% Conservative: 21
Best Local Similarity: 24.4% Mismatches: 35
Query Match: 8.5% Indels: 34
DB: 6 Gaps: 4

US-10-718-311-4 (1-165) x US-10-953-349-34874 (1-1406)
Qy 51 IleArgGluGlyPheValGluGlnAsnGluLeuPro-----GluGlu 64
Db 906 CTGAGGTCGAGCAGCGTGAGCTTCTCCAGCTTCCGAGTCTCCGCGGGATGCGAGGAGAG 847
Qy 65 LeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeuLeuCysAlaAsp 84
Db 846 ATGCCGCTGTTGAGGAGG----- 829
Qy 85 GlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeuSerGlyProGlu 104
Db 828 -----ATGAGCTCGTTGAGGTGCGCGCC 805
Qy 105 LeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPheThrSerSerThr 124
Db 804 ATGTCCCGATGTCTCGGCGAGCGGCTCACCCGCGGAGGTTTC---GCCAGCAGC 748
Qy 125 LeuThrArgAspPheIleGluLeuGlyArgAspAlaGlyLeuTrpGlyArgSerArg 144
Db 747 ATCACC---GACCGCGTCGAGTTGGTGAAGCTGCGCGCAGCGTGAACCTCGAAGTGGTTG 691
Qy 145 LeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuProAlaSerPro 163
```

```
Db 690 TTGTTGATGAAAGCGGTGATCTTCTTGTGGAAGATGCCCGCGGCACCTGCGCG 634
RESULT 9
; Sequence 665, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 665
; LENGTH: 3023
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-665
Alignment Scores:
Pred. No.: 50.6 Length: 3023
Score: 72.00 Matches: 35
Percent Similarity: 38.6% Conservatives: 16
Best Local Similarity: 26.5% Mismatches: 53
Query Match: 8.4% Indels: 28
DB: 7 Gaps: 5
US-10-718-311-4 (1-165) x US-11-293-697-665 (1-3023)
QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 448 TTAGAGACCCACATATAGTAGCTGACCACTGAGAGGTGTGATGAATTAACAGGG 507
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 508 CTCCATCTCTAAAAAACAACGCCACTTGCTG-----CACCTTAGAGAACGATGGGAG 558
QY 41 GlnGln-----GlyLysThrValSerVal 48
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 559 CAGCAGGTGTGCGCAGCAGATGCGAAACCTGCGCGCAAGAGGAAAGTGAACCCAG 618
QY 49 ThrMetIleArgGluGlyPheValGluGlnAsnGluIleProGluGluLeuPro----- 66
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 619 GCACCTCAGCTGAGCCATTCTCAGGGGACTACATCCTGAAGAGAACTGGCAGG 678
QY 67 -----LeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeuLeuCysAlaAsp 84
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 679 AAAAGGCGAGAGCCCAAGGCAACAGAGCTGTCGGAAGAGTCTCTTAAACCCAGTGAC 738
QY 85 GlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeuSerGlyProGlu 104
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 739 AATGAACAAGGCTTG-----CTGTGTCTCCGGCTCTCCGCC----- 777
QY 105 LeuAlaLeuGlnLysLeuGlyLysThrProLeuGly 116
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 778 -----ATGAGAGTCTTTCATCCACAGTGAGGC 807
RESULT 10
US-10-524-433-1
; Sequence 1, Application US/10524433
; Publication No. US20060099676A1
; GENERAL INFORMATION:
; APPLICANT: Functional Genetics, Inc.
; TITLE OF INVENTION: MAMMALIAN GENES INVOLVED IN RAPAMYCIN
; FILE REFERENCE: 10784-023-228
; CURRENT APPLICATION NUMBER: US/10/524,433
; CURRENT FILING DATE: 2005-02-15
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
US-11-293-697-200/c
; Sequence 200, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 200
; LENGTH: 2490
; TYPE: DNA
```

```
; SEQ ID NO 1
; LENGTH: 966
; TYPE: DNA
; ORGANISM: murine
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 548, 564, 567, 583, 603, 610, 636, 647, 648, 651, 654, 656,
; LOCATION: 661, 667, 669, 675, 691, 710, 713, 719, 732, 746, 749, 754,
; LOCATION: 761, 771, 777, 781, 792, 795, 802, 808, 812, 813, 824, 833,
; LOCATION: 841, 842, 856, 861, 866, 871, 873, 874, 875, 882, 887
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 897, 902, 905, 910, 915, 923, 927, 928, 935, 938, 948, 949,
; LOCATION: 965, 966
; OTHER INFORMATION: n = A,T,C or G
US-10-524-433-1
Alignment Scores:
Pred. No.: 11.2 Length: 966
Score: 71.50 Matches: 33
Percent Similarity: 40.8% Conservatives: 18
Best Local Similarity: 26.4% Mismatches: 56
Query Match: 8.4% Indels: 19
DB: 6 Gaps: 4
US-10-718-311-4 (1-165) x US-10-524-433-1 (1-966)
QY 39 PheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGln 58
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 385 TTAAACACACAGGAAAG---TACACAGCCTCCCGCTCAGGCGGAGCCAGTGAAGT 441
QY 59 AsnGluIleProGluGluLeuProLysGluSerArgTyrTrpLeuArgGlu 78
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 442 ACCCACCCTTCCTTGCATCTGCCACACCTGGGCACAGGAGCCTGCTGCCGCGCTGG 501
QY 79 IleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSer 98
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 502 GCTTTGG--TGC-----TGCTCTGTACTCCACCTTGTCTGTGTGGTGGCC 544
QY 99 ThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyr 118
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 545 TCAGGAAAGGT---CATCTCTNAAGAAACAAGCCACCTTCTCCCTCCCTGGCATG 601
QY 119 LeuPheThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeu 138
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 602 CMTTTTTCGCCACAGCCCGTGTGTGCCAACANAAATTTTGGCNCNANGANGNAGGN 661
QY 139 TrpGlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPhe 158
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 662 TGGGANAAACCCNAAACTA-----TTGTGG 688
QY 159 LeuProAlaSerPro 163
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 689 CTNCCCTGTGTGCCCC 703
RESULT 11
US-11-293-697-200/c
; Sequence 200, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 200
; LENGTH: 2490
; TYPE: DNA
```



Qy	120	PheThrSerSerThrLeu-ThrArgAspPheIleGluLeuGlyArgAspAlaGlyLeuTr	139
Db	832	GAACAATCTCTCTCTACACAAG--TTTACTCAGGTTTGACAT-----CTTTT	879
Qy	139	pGlyArgArgSerArgLeuArgLeuSerGlyIysProLeuLeuLeuThr	155
Db	880	TGGGCGAAAGAACCGAAGTAGATGCTTCCTCCCTCCGTCCTTGGTTACC	928

RESULT 15  
US-10-523-014-2/c  
; Sequence 2, Application US/10523014  
; Publication No. US2006094101A1  
; GENERAL INFORMATION:

```

; APPLICANT: Lin, Lin-Ling
;
; APPLICANT: Yannoni, Yvonne
;
; TITLE OF INVENTION: MK2 Interacting Proteins
;
; FILE REFERENCE: 08702-0097-00304
;
; CURRENT APPLICATION NUMBER: US/10/523,014
;
; CURRENT FILING DATE: 2005-02-01
;
; PRIOR APPLICATION NUMBER: USSN 60/400,044
;
; PRIOR FILING DATE: 2002-08-02
;
; NUMBER OF SEQ ID NOS: 6
;
; SOFTWARE: PatentIn version 3.1
;
; SEQ ID NO 2
;
; LENGTH: 2555
;
; TYPE: DNA
;
; ORGANISM: Homo sapiens
;
; US-10-523-014-2

```

Qy	78	GlulleLeuLeuCysAlaAaspGlyclulProTrp	-----	88
		:::		
Db	896	CAGTGCCTGTGGCTGCTCTTGGTATCCTTGGTAA	GGTGGCAGACGTGGCTTTTGTGG	837
Qy	89	-LeuAlaGlyAtGThr-	-----	93
		:::		
Db	836	CCGACGGCGTGTGGTTCGCAGCTCCTGCTCCTTGC	AGCTTGTCCGCTGTAGTGGAA	777

Qy	94	-----ValValProValSerThrLeuSerGlyProGlu	104
Db	776	AAGTCCACCCGTTTGGTGCATCCACCGCTTGTACCTTTTGCACAAGCATCGAAGAA	717
Qy	105	Leu-----AlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrrLeu-----	119
Db	716	CGCGTTGGACGCTTGAACCTTATAGGCAAGTCCACCCGGCCACAGAGCTCACATTGAG	657
Qy	120	PhethrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrp	139
Db	656	TTTGAGGGGAGCACCCCTCTTTTGGATTCTTGCAGATAGGG-----	615
Qy	140	GlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeu	159
Db	614	-----CTCCCTCCATCTCCGAGTCAGTGGTGGTGTGTGATCTCTGCTGTGGAGTTTCTC	561
Qy	160	ProAlaSerProLeu	164
Db	560	AGGCAGGAACCCCTG	546

Search completed: May 30, 2006, 05:54:18  
Job time : 39.9701 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 30, 2006, 01:43:54 ; Search time 4259.96 Seconds  
(without alignments)

3248.869 Million cell updates/sec

Title: US-10-718-311-4

Perfect score: 854

Sequence: 1 MSHPALQALRYLRYCKEIPA.....RLSGKPLLLTEFLPASPLY 165

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 48236798 seqs, 27959665780 residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/abs/ABSSWEB spool/US10718311/runat\_26052006\_164919\_13483/app\_query.fasta\_1  
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abes02h  
-USER=US10718311 @CGN 1.1 7986 @runat\_26052006\_164919\_13483 -NCPU=6 -ICPU=3  
-NO\_WMAP -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: \*  
1: gb\_est1: \*  
2: gb\_est3: \*  
3: gb\_est4: \*  
4: gb\_est5: \*  
5: gb\_est6: \*  
6: gb\_hic: \*  
7: gb\_est2: \*  
8: gb\_est7: \*  
9: gb\_est8: \*  
10: gb\_est9: \*  
11: gb\_ges1: \*  
12: gb\_ges2: \*  
13: gb\_ges3: \*  
14: gb\_ges4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	618	72.4	832	13	CL666350 PRI0152b
2	508	59.5	828	13	CL687382 PRI0146c
3	262	30.7	536	14	AG192190 Pan trogl
4	236	27.6	533	14	AG192116 Pan trogl

#### ALIGNMENTS

RESULT 1  
CL666350/c

LOCUS  
DEFINITION

CL666350 PRI0152b\_A03 - PRI0152b.B21 (832) Mixed stage fosmid library of P. pacificus var. California Pristionchus pacificus genomic, genomic survey sequence.

ACCESSION  
VERSION

CL666350 GI:50159025

KEYWORDS  
SOURCE

GSS.  
Pristionchus pacificus

ORGANISM

Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.

REFERENCE  
AUTHORS

Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J. AppADB: an AcedB database for the nematode satellite organism Pristionchus pacificus

TITLE

Nucleic Acids Res. 32 (1), D421-D422 (2004)

JOURNAL  
PUBMED

14681447

COMMENT

Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498  
Email: raif.sommer@tuebingen.mpg.de  
This library was generated at Caltech, Pasadena, USA and end  
sequenced at Vancouver, Canada.

CC048367 01S0415-0  
CL650764 PRI0110b  
CL688660 PRI014a\_D  
AG192305 Pan trogl  
BZ551713 pac81-60  
BZ564576 pac82-164  
AJ863670 Ralstonia  
BZ578851 msh2\_6034  
CL696699 PRI019d\_B  
BG799182 fp31c10\_Y  
BJ932163 BJ932163  
DY782699 ASXB560.b  
DY936666 III-952-7  
CO384679 AGENCOURT  
AV850332 Magnaport  
BM870249 mgns009XO  
CE278012 tigr-g88-  
CX288046 C01016B12  
C0161709 FLD1\_30\_D  
DA934706 DA934706  
BQ642631 AGENCOURT  
BQ650781 AGENCOURT  
CK209417 FGAS02118  
BJ578705 BJ578705  
BJ566157 BJ566157  
CL393085 ZMMBB019  
CV120546 AGENCOURT  
BU513771 AGENCOURT  
CF513983 Cabud0007  
BF543156 UI-R-AF1-  
BM265186 fw63f10\_Y  
BJ974078 BJ974078  
BZ578446 msh2\_5870  
CV242171 WS02514\_B  
CL106163 ISB1-46F6  
CG472449 ZMMBB028  
AW236750 xms4e06.x  
AW058694 f650b08.Y  
DA275761 DA275761  
DB224823 DB224823  
BC056456 Mus muscu

Seq primer: T7  
Class: fosmid ends.

This library was generated at Caltech, Pasadena, USA and end  
sequenced at Vancouver, Canada.  
Seq primer: T7  
Class: fosmid ends.

#### FEATURES

source  
1. 832  
/organism="Pristionchus pacificus"  
/mol\_type="genomic DNA"  
/strain="California"  
/db\_xref="taxon:54126"  
/clone\_lib="Mixed stage fosmid library of P. pacificus  
var. California"  
/note="Vector: pEpifos-5 Fosmid vector"

Location/Qualifiers  
1. 828  
/organism="Pristionchus pacificus"  
/mol\_type="genomic DNA"  
/strain="California"  
/db\_xref="taxon:54126"  
/clone\_lib="Mixed stage fosmid library of P. pacificus  
var. California"  
/note="Vector: pEpifos-5 Fosmid vector"

#### ORIGIN

Alignment Scores:  
Pred. No.: 9,496-58 Length: 832  
Score: 618.00 Matches: 120  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 72.4% Indels: 0  
DB: 13 Gaps: 0

US-10-718-311-4 (1-165) x CL666350 (1-832)

Qy 46 ValSerValThrMetIleArgGluGlyPheValGluGlnAsnGluIleProGluGluLeu 65  
Db 832 GTAACGGTACGATGATCCGAGAGGGTTTGTGAGCAGATGAATCCCGAGACTG 773  
Qy 66 ProLeuLeuProLyGluSerArgTyrTrpLeuArgGluIleLeuLeuCyAAlaAaspGly 85  
Db 772 CGCGTGTCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTTGTATGTCCGATG 713  
Qy 86 GluProTrpLeuAlaGlyArgThrValProValSerThrLeuSerGlyProGluLeu 105  
Db 712 GAACCGTGGCTTGGCGGCTGATCCGCTGTTCTGTGTCAACGTTAAGCGGCGGAGCTG 653  
Qy 106 AlalaGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPheThrSerThrLeu 125  
Db 652 CGGTACAAATAATGGGTAAACGCCGTTAGACGCTATCTGTTACATCATCGACATTA 593  
Qy 126 ThrArgAaspPheIleGluIleGlyArgAaspAlaGlyLeuTrpGlyArgSerArgLeu 145  
Db 592 ACCGGGACTTATTGAGATAGCGCTGATCCGGGCTGTGGGGCGGACGTTCCCGCTG 533  
Qy 146 ArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuProAlaSerProLeuTyr 165  
Db 532 CGATTACCGGTAAACCGCTGTTGCTAACAGAACTGTTTACCGGCGTCACCGTTGTAC 473

RESULT 2  
CL687382  
LOCUS  
DEFINITION  
PR10146c\_H03.2 - PR10146c.BR. (828) Mixed stage fosmid library of P.  
pacificus var. California Pristionchus pacificus genomic, genomic  
survey sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

CL687382.1 GI:50196135  
GSS.  
Pristionchus pacificus  
Pristionchus pacificus  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

REFERENCE  
AUTHORS  
TITLE

Srinivasan, J., Otto, G.W., Kahlow, U., Geisler, R. and Sommer, R.J.  
AppADB: an AcedB database for the nematode satellite organism  
Pristionchus pacificus  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

JOURNAL  
PUBMED  
COMMENT

Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498  
Email: ralf.sommer@tuebingen.mpg.de

AGI92190  
LOCUS  
DEFINITION  
Pan troglodytes DNA, clone: RP43-068J06.T7, genomic survey  
sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AGI92190.1 GI:45224366  
GSS.  
Pan troglodytes (chimpanzee)  
Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Pan.

REFERENCE  
AUTHORS

Park, H., Kim, Y., Kim, S., Han, Y., Woo, T., Park, K., Eun, C.J.,  
Hoon, S.T., Chu, M., Kim, H., Joo, S., Kim, C., Song, W. and Yoo, H.

BAC end sequences of Library RP-43  
Unpublished  
2 (bases 1 to 536)

REFERENCE  
AUTHORS  
TITLE

Park, H., Kim, Y., Kim, S., Han, Y., Woo, T., Park, K., Eun, C.J.,  
Hoon, S.T., Chu, M., Kim, H., Joo, S., Kim, C., Song, W. and Yoo, H.  
Direct Submission  
Submitted (07-JAN-2002) Hong-Seog Park, Korea Research Institute of  
Bioscience and Biotechnology (KRIBB), Genome Research Center (GRC);  
52, Oun-dong, Yusong-gu, Daejeon 305-333, Korea  
(E-mail: redstone@mail.krribb.re.kr, URL: http://phs.grc.krribb.re.kr/,  
Tel: 82-42-866-7181, Fax: 82-42-860-4409)

COMMENT

Clones are derived from the chimpanzee BAC library RP-43 This BAC



end was generated during the R&D process and may have higher chance of clone tracking errors.

## PRIMERS

Sequencing: T7

LIBRARY  
Vector : pBACe3.6  
R.Site 1 : EcoRI  
R.Site 2 : EcoRI

## FEATURES

source

Location/Qualifiers  
1..536  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/clone="RP43-068J06.T7"  
/sex="male"  
/cell\_type="lymphocytes"  
/clone\_lib="RP-43 Chimpanzee Male BAC Library"

## ORIGIN

Alignment Scores:  
Pred. No.: 1.67e-18 Length: 536  
Score: 262.00 Matches: 51  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 30.7% Indels: 0  
DB: 14 Gaps: 0

US-10-718-311-4 (1-165) x AG192190 (1-536)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20  
|||||  
Db 384 ATGTACACCCCGCTTAACGCAACTGCTGGCGTCTATTGTAAGAGATCCCTGCC 443  
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40  
|||||  
Db 444 CTGGATCCCACTGCTGCACCTGGCTGTGGAGGATTCATGACAAACGTTTGA 503  
QY 41 GlnGlnGlyLysThrValSerValThrMetIle 51  
|||||  
Db 504 CAGCAGGAGAAACGGTAAAGCGTGACGATGATC 536

RESULT 4  
AG192116  
LOCUS  
DEFINITION AG192116 533 bp DNA linear GSS 06-MAR-2004  
Pan troglodytes DNA, clone: RP43-068H07.T7, genomic survey  
sequence.  
ACCESSION AG192116  
VERSION AG192116.1 GI:45224292  
KEYWORDS GSS.  
SOURCE Pan troglodytes (Chimpanzee)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Pan.

## REFERENCE

## AUTHORS

Park, H., Kim, Y., Kim, S., Han, Y., Woo, T., Park, K., Eun, C.J.,  
Hoon, S.T., Chu, M., Kim, H., Joo, S., Kim, C., Song, W. and Yoo, H.

## TITLE

BAC end sequences of Library RP-43

## JOURNAL

## REFERENCE

Unpublished

## AUTHORS

2 (bases 1 to 533)

## TITLE

Direct Submission

## JOURNAL

Submitted (07-JAN-2002) Hong-Seop Park, Korea Research Institute of

Bioscience and Biotechnology (KRIBB), Genome Research Center (GRC);

52, Oun-dong, Yuseong-gu, Daejeon 305-333, Korea

(E-mail: redstone@mail.krribb.re.kr, URL: http://phs.grc.krribb.re.kr/,

Tel: 82-42-866-7181, Fax: 82-42-860-4409)

Clones are derived from the chimpanzee BAC library RP-43 This BAC

end was generated during the R&amp;D process and may have higher chance

of clone tracking errors.

## PRIMERS

Sequencing: T7

LIBRARY

Vector : pBACe3.6  
R.Site 1 : EcoRI  
R.Site 2 : EcoRI

## FEATURES

source

Location/Qualifiers  
1..533  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/clone="RP43-068H07.T7"  
/sex="male"  
/cell\_type="lymphocytes"  
/clone\_lib="RP-43 Chimpanzee Male BAC Library"

## ORIGIN

Alignment Scores:  
Pred. No.: 1.27e-15 Length: 533  
Score: 236.00 Matches: 47  
Percent Similarity: 92.2% Conservative: 0  
Best Local Similarity: 92.2% Mismatches: 4  
Query Match: 27.6% Indels: 0  
DB: 14 Gaps: 0

US-10-718-311-4 (1-165) x AG192116 (1-533)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 20  
|||||  
Db 380 ATGTACACCCCGCTTAACGCAACTGCTGGCGTCTATTGTAAGAGATCCCTGCC 439  
QY 21 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 40  
|||||  
Db 440 CTGGATCCCACTGCTGCACCTGGCTGTGGAGGATTCATGACAAACCTTTTGA 499  
QY 41 GlnGlnGlyLysThrValSerValThrMetIle 51  
|||||  
Db 500 CAGCAGGAGAAACGGTATTCGTGACGATGATC 532

## RESULT 5

## CC048367

## LOCUS

## DEFINITION

CC048367

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Zea mays

Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE

## AUTHORS

## TITLE

Latehaw, S., Tan, B.-C., Settles, A.M. and McCarty, D.R.

Sequence tagged transposon insertions from the UniformMu maize

population

Unpublished (2003)

## JOURNAL

## COMMENT

Contact: Donald R. McCarty

Plant Molecular and Cellular Biology Program

University of Florida

PO 110690 Gainesville, FL 32611-0690, USA

Tel: 352-392-1928 x322

Email: drmc@ufl.edu

Sequence flanking probable Mu insertion site in UniformMu line:

01S-415-4

Class: transposon insertion site.

Location/Qualifiers

1..515

/organism="Zea mays"

/mol\_type="genomic DNA"

/strains="W22 (ACR, bz1-m9)"

/cultivar="UniformMu"

/db\_xref="taxon:4577"

/clone="01S-415-4-7to12-E12"

/clone\_lib="UniformMu MuTAIL Library"



```

Db      ||||||| 785 CCTGGATCCGCAACTGCTCGACTGGCTGTGTGCTGGAGGATTCCATGACAAA 836
RESULT 8
AG192305
LOCUS   440 bp DNA linear GSS 06-MAR-2004
DEFINITION Pan troglodytes DNA, clone: RP43-068M07.T7, genomic survey
sequence.
ACCESSION AG192305
VERSION   AG192305.1 GI:45224481
KEYWORDS  GSS
SOURCE    Pan troglodytes (chimpanzee)
ORGANISM  Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Pan.
REFERENCE
1 Park, H., Kim, Y., Kim, S., Han, Y., Woo, T., Park, K., Eun, C.J.,
Hoon, S.T., Chu, M., Kim, H., Joo, S., Kim, C., Song, W. and Yoo, H.
BAC end sequences of Library RP-43
Unpublished
2 (bases 1 to 440)
Park, H., Kim, Y., Kim, S., Han, Y., Woo, T., Park, K., Eun, C.J.,
Hoon, S.T., Chu, M., Kim, H., Joo, S., Kim, C., Song, W. and Yoo, H.
Direct Submission
Submitted (07-JAN-2002) Hong-Seog Park, Korea Research Institute of
Bioscience and Biotechnology (KRIBB), Genome Research Center (GRC);
52, Oun-dong, Yusong-gu, Daejeon 305-333, Korea
(E-mail:redstone@mail.krrib.re.kr, URL:http://phs.grc.krrib.re.kr/,
Tel:82-42-866-7181, Fax:82-42-860-4409)
Clones are derived from the chimpanzee BAC library RP-43 This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.
PRIMERS
Sequencing: T7
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES             Location/Qualifiers
     1..440
     /organism="Pan troglodytes"
     /mol_type="genomic DNA"
     /db_xref="taxon:9598"
     /clone="RP43-068M07.T7"
     /sex="male"
     /cell_type="lymphocytes"
     /clone_lib="RP-43 Chimpanzee Male BAC Library"
ORIGIN
Alignment Scores:
Pred. No.:      2,64e-07      Length:      440
Score:          160.00      Matches:      30
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches:  0
Query Match:     18.7%      Indels:       0
DB:              14      Gaps:          0

US-10-718-311-4 (1-165) x AG192305 (1-440)

QY      1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysIysGluIleProAla 20
Db      ||||||| 349 ATGTACACCCCGCGTTAAGCAACTGCTGCGCTATTGTTAAAGAGATCCCTGCC 408
QY      21 LeuAppProGlnLeuLeuAspTrpLeuLeu 30
Db      ||||||| 409 CTGGATCCGCAACTGCTCGACTGGCTGTTG 438

RESULT 9
BZ551713
LOCUS   1110 bp DNA linear GSS 17-DEC-2002
DEFINITION pacs1-60_3287.x1 pacs1-60 Pseudomonas aeruginosa genomic clone
pacs1-60_3287, genomic survey sequence.

```

```

ACCESSION BZ551713
VERSION   BZ551713.1 GI:27155321
KEYWORDS  GSS
SOURCE    Pseudomonas aeruginosa
ORGANISM  Pseudomonas aeruginosa
REFERENCE 1 (bases 1 to 1110)
AUTHORS   Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE     Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL   J. Bacteriol. (2002) In press
COMMENT   Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES             Location/Qualifiers
     1..1110
     /organism="Pseudomonas aeruginosa"
     /mol_type="genomic DNA"
     /strain="1-60"
     /db_xref="taxon:287"
     /clone="pacs1-60_3287"
     /note_lib="pacs1-60"
     /note="clinical isolate 1-60 Whole genomic shotgun
     library."
ORIGIN
Alignment Scores:
Pred. No.:      1.28e-05      Length:      1110
Score:          150.00      Matches:      45
Percent Similarity: 37.7%      Conservative: 12
Best Local Similarity: 29.8%      Mismatches:  56
Query Match:     17.6%      Indels:       38
DB:              11      Gaps:          3

US-10-718-311-4 (1-165) x BZ551713 (1-1110)

QY      21 LeuAppProGlnLeuLeuAspTrpLeuLeuLeuLeuLeuGluAspSerMetThrLysArgPheGlu 40
Db      ||||||| 144 GTCGAACCCCTGCTGGAGGCTGGCAGACCCTGCGGACGAC-----GAA 188
QY      41 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60
Db      ||||||| 189 TGCCAGGGGCTCGACGTC----- 206
QY      61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db      ||||||| 207 -----CCTACCGCAGTAGCGCTGGCTGCCGAGGTCTAC 242
QY      81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100
Db      ||||||| 243 CTGCATGGCCACGACCGTCCCTGGGTGTTCGCCCGCAGCGTGGCGCGCGCCCTG 302
QY      101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db      ||||||| 303 GAAGGCTCGGGCTTACGCTGGCGTGTCTGGCACCGCTCGTGGCGAGTTGCTGTTTC 362
QY      121 ThrSerSerThrLeuThrArgAspPheIleGluIle----- 132
Db      ||||||| 363 AGCAGACGCGCTTCGAGCGCGGGCCCATCGAAGTCTGCCGCTATCCGCGCGCGCTCTG 422
QY      133 -----GlyArgAspAlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGly 149
Db      ||||||| 423 CCCGCGGAGTCCCGCGGAGGTTCTTGGGGCCGCTCGCTACGGTTTCCCGCGCGCG 482
QY      150 LysProLeuLeuLeuThrGluLeuPheLeuPro 160
Db      ||||||| ::::|::: ::::|:::

```

```

Db      483 CTCGGGTCCTGGTGGCGGAGTGACTACCG 515

RESULT 10
BZ564576/c
LOCUS   BZ564576               862 bp    DNA        linear    GSS 17-DEC-2002
DEFINITION   pacs2-164_4819.v2 pacs2-164 Pseudomonas aeruginosa genomic clone
pacs2-164_4819, genomic survey sequence.
ACCESSION   BZ564576
VERSION     BZ564576.1  GI:27190626
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
ORGANISM    Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE   1 (bases 1 to 862)
AUTHORS     Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE       Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL     J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel.: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES    Location/Qualifiers
             source
               1. 862
               /organism="Pseudomonas aeruginosa"
               /mol_type="genomic DNA"
               /strain="2-164"
               /db_xref="taxon:287"
               /clone="pacs2-164_4819"
               /clone_lib="pacs2-164"
               /note="clinical isolate 2-164 Whole genomic shotgun
               library."
ORIGIN
Alignment Scores:
Pred. No.:      0.000191      Length:      862
Score:          138.00      Matches:     45
Percent Similarity: 37.3%      Conservative: 11
Best Local Similarity: 30.0%      Mismatches:  56
Query Match:    16.2%      Indels:      38
DB:             11      Gaps:         3

US-10-718-311-4 (1-165) x BZ564576 (1-862)

Qy      22 AspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGluGln 41
      : : : : :
Db      662 AACCCCTGCCTGGAGGCGCTGGCGAGACCTGGCGCGAGC-----GAATGC 618
      : : : : :

Qy      42 GlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGluIle 61
      : : : : :
Db      617 CAGGGGCTGCAGCTC----- 603

Qy      62 ProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeuLeu 81
      : : : : :
Db      602 -----CCTACCGCGAGTAGCGGCTGGTCCGCGAGGCTACCTG 564

Qy      82 CysAlaAspGlyLupProTrpLeuAlaGlyArgThrValValProValSerThrLeuSer 101
      : : : : :
Db      563 CATGCCACGACCGTCCCTCGAGTGTTCGCCCGCAGCGTGGCGGCGCGAGCGCCCTGGAA 504

Qy      102 GlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPheThr 121
      : : : : :
Db      503 GGCTCGGGCTTCACCTGGCGGTGTTCGGACCCCGCTCGTGGGCGAGTGTCTGTTCAGC 444

Qy      122 SerSerThrLeuThrArgAspPheIleGluIle----- 132
      : : : : :
Db      443 GACAGCGCTTCGAGCGCGGGGCCCATCGAAGTCTCGCGCTATCCGCGCGCGGTCTGCC 384

```

```

Qy      133 -----GlyArgAspAlaGlyLeuTrpGlyArgArgSerArgLeuArgLeuSerGlyLys 150
      : : : : :
Db      383 GCCGAGGTCCCGCGCGAGGGTCTCTGGGGCGCTCGCTCACGGTTTTCGCCGCGCGCTC 324
      : : : : :

Qy      151 ProLeuLeuLeuThrGluLeuPheLeuPro 160
      : : : : :
Db      323 GGGGTGCTGGTGGCGGAGGTGTACTACCG 294

RESULT 11
AJ863670/c
LOCUS   AJ863670               739 bp    DNA        linear    GSS 30-NOV-2005
DEFINITION   Ralstonia solanacearum GSS, clone V789R, genomic survey sequence.
ACCESSION   AJ863670
VERSION     AJ863670.1  GI:82937082
KEYWORDS    GSS; genome survey sequence.
SOURCE      Ralstonia solanacearum
ORGANISM    Ralstonia solanacearum
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Burkholderiaceae; Ralstonia.
REFERENCE   1 (bases 1 to 739)
AUTHORS     Munoz-Soriano, V., Arahal, D.R., Terol, J., Buades, C., Perez-Perez, A.,
Llop, P., Belmonte, U.C.F., Lopez, M. and Perez-Alonso, M.
TITLE       Random genome sequencing of Ralstonia solanacearum strain IVIA 1602
and comparative analysis with strain GM1000
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 739)
AUTHORS     Munoz-Soriano, V.
TITLE       Direct Submission
JOURNAL     Submitted (18-NOV-2004) Genetica, Universidad de Valencia, C/ Dr.
Moliner, 50, Burjassot, Valencia 46100, SPAIN
FEATURES    Location/Qualifiers
             source
               1. 739
               /organism="Ralstonia solanacearum"
               /mol_type="genomic DNA"
               /strain="IVIA 1602"
               /db_xref="taxon:305"
               /clone="V789R"
ORIGIN
Alignment Scores:
Pred. No.:      0.0254      Length:      739
Score:          118.00      Matches:     42
Percent Similarity: 38.9%      Conservative: 23
Best Local Similarity: 25.1%      Mismatches:  64
Query Match:    13.8%      Indels:      38
DB:             14      Gaps:         3

US-10-718-311-4 (1-165) x AJ863670 (1-739)

Qy      25 LeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGluGlnGlnGlyLys 44
      : : : : :
Db      599 CTGTTGATGTTGGTGAAGATCTTGAATTCGTGACGCGCGGCTGCGTGGCGTTCCTCC 540
      : : : : :

Qy      45 ThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGluIleProGluGlu 64
      : : : : :
Db      539 TCGTTCGCGTCAACTGCTCGGTGAG-----CAGCGGACC 504

Qy      65 LeuProLeu-----LeuProLysGluSerArgTyrTrpLeu 76
      : : : : :
Db      503 ATGCGCTGAGCAGCAGAGTGGCGTGCCTGGGCTTGCCCGTGGCGGAGAGCTGGCG 444

Qy      77 ArgGluIleLeuLeuCysAlaAspGlyLupProTrpLeuAlaGlyArgThrValValPro 96
      : : : : :
Db      443 CGCAAGTCTCTGCTGATCTGCGACGAAACCCCGGTGGTCTACGCACACACCATGTCGAT 384
      : : : : :

Qy      97 ValSerThrSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGly 116
      : : : : :
Db      383 CCGCGCAGCTGGCGGCGGACTGGCCGTTCTCTGAAGCGCTGGGACCCAGCGCTCGGG 324
      : : : : :

Qy      117 ArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGlu----- 131
      : : : : :
Db      323 CATCGCTGTTTCGCGGATCCGCGGTTGCACGGGCGCTTCGAGTTTGTGCCAGCTGGAT 264

```

QY 132 -----1leGlyArg 134  
 Db 263 GTCCGGATCCGCTGGTCAGCGGGCCGTCGGCGCTGGCGCGCAGCGCGCTGGCGGC 204  
 QY 135 AsPAlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeu 154  
 Db 203 ATGCGCGGTCGCGCGCAGCCGCTTCGGTATTTCGCGCGGTGCCAGCGCGATGCTGGTG 144  
 QY 155 ThrGluLeuPheLeuProAla 161  
 Db 143 ACGAAAGTGTTCCTGCCTGCA 123

RESULT 12  
 LOCUS BZ578851 1141 bp DNA linear GSS 17-DEC-2002  
 DEFINITION msh2\_6034.x1 msh Pseudomonas aeruginosa genomic clone msh2\_6034,  
 genomic survey sequence.

ACCESSION BZ578851  
 VERSION BZ578851  
 KEYWORDS GSS.

SOURCE Pseudomonas aeruginosa  
 ORGANISM Pseudomonas aeruginosa

REFERENCE Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.

AUTHORS 1 (bases 1 to 1141)  
 Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
 Burns,J.L., Kaul,R. and Olsen,M.V.

TITLE Whole-Genome-Sequence variation among multiple isolates of  
 Pseudomonas aeruginosa library

JOURNAL J. Bacteriol. (2002) in press  
 COMMENT Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Class: shotgun.

FEATURES  
 source  
 1..1141  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="M5H"  
 /db\_xref="taxon:287"  
 /clone="msh2\_6034"  
 /clone\_lib="msh"  
 /note="Environmental isolate. Whole genomic shotgun  
 library."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.31 Length: 1141  
 Score: 105.00 Matches: 44  
 Percent Similarity: 36.2% Conservative: 24  
 Best Local Similarity: 23.4% Mismatches: 70  
 Query Match: 12.3% Indels: 50  
 DB: 12 Gaps: 6

US-10-718-311-4 (1-165) x BZ578851 (1-1141)

QY 17 GluIleProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThr 36  
 Db 466 CAACGTCCAGGCTCTCCCCACCGCTCTCGACTGCTTTCGACGAAGGGTCGTGACC 525

QY 37 LysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheVal 56  
 Db 526 CGCGCGC-----CTGACCGTCTCTCGCGCGCGCGGTTC 558

QY 57 GluGlnAsnGluIleProGluGluLeuProLeuLeu----- 68  
 Db 559 CGCGTCAACCCCTGTGGAAGGCTGGACAGCCCTGCGCAGCAGCAATGCCAGGGGCTC 618

QY 69 -----ProLysGluSerArgTrpTrpLeuArgGluIleLeuLeuCysAlaAspGlyGlu 86  
 Db 619 GAGGTCCCTCCCGCAGTAGCGGTCGGGTCGGAGGTTACTGTGATGCCACGACCGT 678  
 QY 87 TrpTrpLeuAlaGlyArgThrValProValSerThrLeuSerGlyPro----- 103  
 Db 679 CCTGGGTGNTGCCCGCAGCGTGGCGCGCGCAGCGCCCTGAAGGCTCGGCTCGACT 738  
 QY 104 -----GluLeuAlaLeuGlnLysLeuGlyLysThrPro 114  
 Db 739 TGGCGCTGCTCGCACCCGCTGTTGGCGAGTTGTTAGGAAAGCCCTTGAGCGCGGC 798  
 QY 115 LeuGlyArgTrpLeuPheThrSerSerThrLeuThrArgAspPheIle-----GluIle 132  
 Db 799 CCATCAAGCTGTGCGTATCCGCGCGCTCTTCCCGCCAGTTCGCCGCGAGGCTCTCTG 858  
 QY 133 GlyArgAspAlaGlyLeuTrpGlyArgArgSerArgLeuArg-----Leu 147  
 Db 859 GGCGCTCCTTACGGTTTCCCGCGCGCTGGTGTGGCGAAGGTACCTACCGGCGCTG 918  
 QY 148 SerGlyLysProLeuLeuLeuThr----- 155  
 Db 919 GGGACAGCCCGAATTGCCGACGATAAACCCCGCTGAATTCGGAAGTCCCAAGTTTG 978  
 QY 156 GluLeuPheLeuProAlaSerPro 163  
 Db 979 AACCTTATCAAGCCTTTGTCCCT 1002

RESULT 13  
 LOCUS CL696699/7c 753 bp DNA linear GSS 10-JUL-2004  
 DEFINITION PRI019d\_B03.2 - PRI019d\_BR (753) Mixed stage fosmid library of P.  
 pacificus var. California Pristionchus pacificus genomic, genomic  
 survey sequence.

ACCESSION CL696699  
 VERSION CL696699.1 GI:50218607  
 KEYWORDS GSS.

SOURCE Pristionchus pacificus  
 ORGANISM Pristionchus pacificus

REFERENCE 1 (bases 1 to 753)  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
 Neodiplogasteridae; Pristionchus.

AUTHORS Appinvasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.  
 TITLE Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.  
 A database for the nematode satellite organism

JOURNAL Pristionchus pacificus  
 PUBLISHED Nucleic Acids Res. 32 (1), D421-D422 (2004)  
 COMMENT 14681447

ORIGIN  
 Alignment Scores:  
 Pred. No.: 9.34 Length: 753  
 Score: 95.00 Matches: 19  
 Percent Similarity: 100.0% Conservative: 0

FEATURES  
 source  
 1..753  
 /organism="Pristionchus pacificus"  
 /mol\_type="genomic DNA"  
 /strain="California"  
 /db\_xref="taxon:54126"  
 /clone\_lib="Mixed stage fosmid library of P. pacificus  
 var. California"  
 /note="Vector: pEpifos-5 Fosmid vector"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 9.34 Length: 753  
 Score: 95.00 Matches: 19  
 Percent Similarity: 100.0% Conservative: 0

FEATURES  
 source  
 1..753  
 /organism="Pristionchus pacificus"  
 /mol\_type="genomic DNA"  
 /strain="California"  
 /db\_xref="taxon:54126"  
 /clone\_lib="Mixed stage fosmid library of P. pacificus  
 var. California"  
 /note="Vector: pEpifos-5 Fosmid vector"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 9.34 Length: 753  
 Score: 95.00 Matches: 19  
 Percent Similarity: 100.0% Conservative: 0

FEATURES  
 source  
 1..753  
 /organism="Pristionchus pacificus"  
 /mol\_type="genomic DNA"  
 /strain="California"  
 /db\_xref="taxon:54126"  
 /clone\_lib="Mixed stage fosmid library of P. pacificus  
 var. California"  
 /note="Vector: pEpifos-5 Fosmid vector"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 9.34 Length: 753  
 Score: 95.00 Matches: 19  
 Percent Similarity: 100.0% Conservative: 0

FEATURES  
 source  
 1..753  
 /organism="Pristionchus pacificus"  
 /mol\_type="genomic DNA"  
 /strain="California"  
 /db\_xref="taxon:54126"  
 /clone\_lib="Mixed stage fosmid library of P. pacificus  
 var. California"  
 /note="Vector: pEpifos-5 Fosmid vector"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 9.34 Length: 753  
 Score: 95.00 Matches: 19  
 Percent Similarity: 100.0% Conservative: 0

FEATURES  
 source  
 1..753  
 /organism="Pristionchus pacificus"  
 /mol\_type="genomic DNA"  
 /strain="California"  
 /db\_xref="taxon:54126"  
 /clone\_lib="Mixed stage fosmid library of P. pacificus  
 var. California"  
 /note="Vector: pEpifos-5 Fosmid vector"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 9.34 Length: 753  
 Score: 95.00 Matches: 19  
 Percent Similarity: 100.0% Conservative: 0

FEATURES  
 source  
 1..753  
 /organism="Pristionchus pacificus"  
 /mol\_type="genomic DNA"  
 /strain="California"  
 /db\_xref="taxon:54126"  
 /clone\_lib="Mixed stage fosmid library of P. pacificus  
 var. California"  
 /note="Vector: pEpifos-5 Fosmid vector"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 9.34 Length: 753  
 Score: 95.00 Matches: 19  
 Percent Similarity: 100.0% Conservative: 0

FEATURES  
 source  
 1..753  
 /organism="Pristionchus pacificus"  
 /mol\_type="genomic DNA"  
 /strain="California"  
 /db\_xref="taxon:54126"  
 /clone\_lib="Mixed stage fosmid library of P. pacificus  
 var. California"  
 /note="Vector: pEpifos-5 Fosmid vector"

```

Best Local Similarity: 100.0%      Mismatches: 0
Query Match:      11.1%      Indels: 0
DB:              13         Gaps: 0

US-10-718-311-4 (1-165) x CL696699 (1-753)

QY      147 LeuSerGlyLysProLeuLeuThrGluLeuPheLeuProAlaSerProLeuTyr 165
DB      751 TTAACGGTAACCGCTGTGCTAAAGAACTGTTTTTACCAGCGCTCACCGTTGTAC 695

RESULT 14
LOCUS   BG799182              591 bp      mRNA      linear      EST 21-MAY-2001
DEFINITION   fpcic10.y1 zebrafish gridded kidney Danio rerio cDNA clone
IMAGE:4744722 5' similar to TR:Q9Y4G8 Q9Y4G8 KIAA0313 PROTEIN. ;,
mRNA sequence.
ACCESSION   BG799182
VERSION     BG799182.1 GI:14163514
KEYWORDS
SOURCE      Danio rerio (zebrafish)
ORGANISM    Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1 (bases 1 to 581)
AUTHORS     Clark, W., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M.,
            Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
            Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
            Persson, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
            Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
            Waterston, R., and Wilson, R.
            WashU Zebrafish EST Project 1998
            Unpublished (1998)
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Leonard Ira Zon DNA Sequencing by:
            Washington University Genome Sequencing Center Clone distribution:
            Genome Systems, St. Louis, Missouri (web address:
            www.genomesystems.com) (email contact: info@genomesystems.com) and
            Research Genetics, Huntsville, Alabama (web address:
            www.resgen.com) (email contact: info@resgen.com) and
            RessourcenzentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            High quality sequence stop: 473.
FEATURES
source
1..581
Location/Qualifiers
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:4744722"
/sex="mixed"
/tissue_type="kidney pooled from 300 wild type adults"
/lab_host="XLOUR"
/clone_lib="zebrafish gridded kidney"
/note="Organ: kidney; Vector: pBK-CMV; Site 1: EcoRI;
Site 2: XhoI; Oligo dT cDNA library constructed from mRNA
pooled from pooled kidney tissue from 300 adult
zebrafish."

ORIGIN
Alignment Scores:      10.8      Length:      581
Pred. No.:      93.00      Matches:      38
Score:      45.4%      Conservative:      26
Percent Similarity:      27.0%      Mismatches:      43
Best Local Similarity:      10.9%      Indels:      34
Query Match:      2         Gaps:      7
DB:
US-10-718-311-4 (1-165) x BG799182 (1-581)

QY      2 SerHisProAlaLeuThr-----GlnLeuArgAlaLeuArgTyrCysGlyGluLeuPro 19
DB      117 AGTAACCCCTGACCTGGCACAGGCGCAGCAAGCATCATCGACTACAGTACACGCCACCA 176

US-10-718-311-4 (1-165) x CL696699 (1-753)

QY      20 AlaLeuAspProGlnLeuLeu-----AspTrpLeuLeu 30
DB      177 GAGTTGCAAGACCAGGTGCTCGCATTTTCAAGCGAGATCAGCAAGCCGCTTACTTACTG 236

QY      31 LeuGluAspSerMetThrLysArg-----PheGluGlnGlnGlyLys 44
DB      237 GCCAATAGAGACACACAGCCAGAGAGAGTCCCAATCTAGCCATAAAGAGAGTTTGGCTTG 296

QY      45 ThrValSer-----ValThrMetIleArgGluGlyPheVal 56
DB      297 TCTCGAGTCCCGAAGCTTTTTCGCTCTGTGAAGTTCAGTCACACAGGAAGGAGTTATC 356

QY      57 GluGlnAsnGluLeuProGluLeuProLeuLeuProLys-----GluSer 72
DB      357 AAGCAGAGACGACTGCCTGATCAACTGTCCAACTGGCTGACAGATCCAACTGAGTGCC 416

QY      73 ArgTyrTrpLeuArgGluLeuLeu-----LeuCysAlaAspGlyGluProTrp 88
DB      417 AGGTACTATCTAAAGAGCAATATGCAAAACCGAGACGCTGTGTCGGATGTGGAAGCACTG 476

QY      89 ---LeuAlaGlyArgThrValProValSerThrLeuSerGlyProGluLeuAlaLeu 107
DB      477 GAGATGCAGAGAGAATCTGTGTCGCCCTCTGTTTGTAGCTCAATGGAGATAGCCCAAT 536

QY      108 Gln 108
DB      537 CAG 539

RESULT 15
LOCUS   BJ932163              666 bp      mRNA      linear      EST 07-OCT-2005
DEFINITION   BJ932163 Daphnia magna cDNA library Daphnia magna cDNA clone
IMAGE:0004_L04.r.3', mRNA sequence.
ACCESSION   BJ932163
VERSION     BJ932163.1 GI:66942524
KEYWORDS
SOURCE      Daphnia magna
ORGANISM    Daphnia magna
            Eukaryota; Metazoa; Arthropoda; Crustacea; Branchiopoda;
            Diplostroica; Cladocera; Anomopoda; Daphniidae; Daphnia.
            watanabe,H., Tatarazako,N., Oda,S., Nishide,H., Uchiyama,I.,
            Morita,M. and Iguchi,T.
            Analysis of expressed sequence tags of the water flea Daphnia magna
            Genome 48 (4), 606-609 (2005)
            16094427
            Contact: Hajime Watanabe
            Okazaki Institute for Bioscience
            National Institute of Natural Sciences
            Higashiyama 5-1, Myodaiji, Okazaki, Aichi, 444-8787, Japan
            Tel: 81-564-59-5237
            Fax: 81-564-59-5236
            Email: watanabe@nibb.ac.jp.
FEATURES
source
1..666
Location/Qualifiers
/organism="Daphnia magna"
/mol_type="mRNA"
/db_xref="taxon:35525"
/clone="WITH001_0004_L04.r"
/sex="female"
/tissue_type="whole body"
/dev stage="adult"
/clone_lib="Daphnia magna cDNA library"

ORIGIN
Alignment Scores:      21.8      Length:      666
Pred. No.:      91.00      Matches:      29
Score:

```

Percent Similarity: 37.5%  
Best Local Similarity: 25.9%  
Query Match: 10.7%  
DB: 2  
Conservative: 13  
Mismatch: 24  
Indels: 46  
Gaps: 5

US-10-718-311-4 (1-165) x BJ932163 (1-666)

```
QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 80
Db 312 GTTCCAGACCATATAGCTTACTT-----TATTGGCTTCGTGCTGCGCTT 356
QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 100
Db 357 TCTTCTTCAACGCGCAGGCGCTGG----- 380
QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120
Db 381 -----AATAAAGTGCCTCATTCCTGTTATTACTTT 410
QY 121 ThrSerSerThrLeuThrArg-----Asp 128
Db 411 TAAGGATTGCACATATCTCGTAGAACGCAACAAACCTGGATGCAAGTTATTTTTC 470
QY 129 PheIleGluIleGlyArgAspAla-----GlyLeuTrpGlyArgArg 142
Db 471 TTCCTTGATTGGCGCGTACGGCGCATGGCGAGTTCGCTTACGGCGCATGTCATTATA 530
QY 143 SerArgLeuArgLeu---SerGlyLysProLeuLeu 153
Db 531 AGTCGATTAGTTTGTCCAGTTTCACGTCACCAATGCTG 566
```

Search completed: May 30, 2006, 04:08:20  
Job time : 4262.96 secs

**THIS PAGE BLANK (USPTO)**



1	684	100.0	684	2	AR428736	Sequence
2	684	100.0	684	2	AR455412	Sequence
3	684	100.0	684	2	AX329368	Sequence
4	684	100.0	684	2	AX382258	Sequence
5	498	72.8	854	2	AR144788	Sequence
6	498	72.8	1498	15	ECOUBIA	
7	498	72.8	2000	2	E11274	
8	498	72.8	2000	15	ECOUBICA	
9	498	72.8	2000	15	ECOUBIPLS	
10	498	72.8	2034	15	ECUBIAC	
11	498	72.8	2348	15	ECUBI	
12	498	72.8	6841	2	AR144789	Sequence
13	498	72.8	110000	15	U00096 <sup>42</sup>	
14	498	72.8	110000	15	AP009048 <sup>42</sup>	Continuation (43 o
15	498	72.8	176195	15	ECOW89	U00006 E. coli chr
16	496.4	72.6	1396	15	DOQ87228	Escherich
17	495	72.4	495	2	A47932	Sequence 1
18	495	72.4	495	2	AR428733	Sequence

```
Db      181  TGGCATATGTCACACCCCGCTTAACGCACTGCGTGCCTGCGCTATTGTAAAGAGATC 240
Qy      241  CTTGCCCTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGATTCCATGACAAAACGT 300
Db      241  CTTGCCCTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGATTCCATGACAAAACGT 300
Qy      301  TTTGAACAGCAGGAGAAAACGGTAAGCGTAGCGATGATCCCGAAGGGTTTGTCTGAGCAG 360
Db      301  TTTGAACAGCAGGAGAAAACGGTAAGCGTAGCGATGATCCCGAAGGGTTTGTCTGAGCAG 360
Qy      361  AATGAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGTTTACGTGAA 420
Db      361  AATGAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGTTTACGTGAA 420
Qy      421  ATTTTGTATTATGTCGCGATGGTGAACCGTGGCTTACCGGTGCTGCTGCTGCTGCTGCA 480
Db      421  ATTTTGTATTATGTCGCGATGGTGAACCGTGGCTTACCGGTGCTGCTGCTGCTGCTGCA 480
Qy      481  ACGTTAAGCGGCGGAGCTGCGCTTACAAAATTTGGGTAAACCGCGTTAGGACGCTAT 540
Db      481  ACGTTAAGCGGCGGAGCTGCGCTTACAAAATTTGGGTAAACCGCGTTAGGACGCTAT 540
Qy      541  CTGTTACATCATCGACATTAAACCGGACCTTTATTGAGATAGGCGCGTGTGCTGCTGCTG 600
Db      541  CTGTTACATCATCGACATTAAACCGGACCTTTATTGAGATAGGCGCGTGTGCTGCTGCTG 600
Qy      601  TGGGGCGACGTTCCCGCTCGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTT 660
Db      601  TGGGGCGACGTTCCCGCTCGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTT 660
Qy      661  TTACCGGCGTCACCGTTGTACTAA 684
Db      661  TTACCGGCGTCACCGTTGTACTAA 684
```

```
RESULT 2
AR455412  AR455412  684 bp  DNA  linear  PAT 20-FEB-2004
LOCUS     Sequence 7 from patent US 6683231.
DEFINITION
ACCESSION AR455412
VERSION   AR455412.1  GI:42689960
KEYWORDS  'Unknown.'
SOURCE    Unknown.
ORGANISM  Unclassified.
REFERENCE 1 (bases 1 to 684)
AUTHORS  Meyer, K., Viitanen, P.V. and Van Dyk, D.E.
TITLE    High level production of p-hydroxybenzoic acid in green plants
JOURNAL  Patent: US 6683231-A 7 27-JAN-2004;
        E. I. du Pont de Nemours and Company; Wilmington, DE
FEATURES  Location/Qualifiers
           source
             1..684
             /organism="unknown"
             /mol_type="genomic DNA"
```

```
ORIGIN
Query Match      100.0%; Score 684; DB 2; Length 684;
Best Local Similarity 100.0%; Pred. No. 6.8e-202;
Matches 684; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  ATGGCTTCTCTGTGCTATTCTTCAGCAGCTGTTGCCACAGCAGCAATGTTACACAAGCT 60
Db      1  ATGGCTTCTCTGTGCTATTCTTCAGCAGCTGTTGCCACAGCAGCAATGTTACACAAGCT 60
Qy      61  AGCATGGTTGCAACCTTTCACCTGGTCTCAAAATCTTCAGGCACCTTTCCCTGTTTACAAAG 120
Db      61  AGCATGGTTGCAACCTTTCACCTGGTCTCAAAATCTTCAGGCACCTTTCCCTGTTTACAAAG 120
Qy      121  CAAAACCTTGACATCATCTTCATGCTAGCAATGGTGAAGAGTTAGCTGATGAGGTG 180
Db      121  CAAAACCTTGACATCATCTTCATGCTAGCAATGGTGAAGAGTTAGCTGATGAGGTG 180
Qy      181  TGGCATATGTCACACCCCGCTTAACGCACTGCGTGCCTGCGCTATTGTAAAGAGATC 240
```

```
Db      181  TGGCATATGTCACACCCCGCTTAACGCACTGCGTGCCTGCGCTATTGTAAAGAGATC 240
Qy      241  CTTGCCCTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGATTCCATGACAAAACGT 300
Db      241  CTTGCCCTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGATTCCATGACAAAACGT 300
Qy      301  TTTGAACAGCAGGAGAAAACGGTAAGCGTAGCGATGATCCCGAAGGGTTTGTCTGAGCAG 360
Db      301  TTTGAACAGCAGGAGAAAACGGTAAGCGTAGCGATGATCCCGAAGGGTTTGTCTGAGCAG 360
Qy      361  AATGAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGTTTACGTGAA 420
Db      361  AATGAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGTTTACGTGAA 420
Qy      421  ATTTTGTATTATGTCGCGATGGTGAACCGTGGCTTACCGGTGCTGCTGCTGCTGCTGCA 480
Db      421  ATTTTGTATTATGTCGCGATGGTGAACCGTGGCTTACCGGTGCTGCTGCTGCTGCTGCA 480
Qy      481  ACGTTAAGCGGCGGAGCTGCGCTTACAAAATTTGGGTAAACCGCGTTAGGACGCTAT 540
Db      481  ACGTTAAGCGGCGGAGCTGCGCTTACAAAATTTGGGTAAACCGCGTTAGGACGCTAT 540
Qy      541  CTGTTACATCATCGACATTAAACCGGACCTTTATTGAGATAGGCGCGTGTGCTTAACAGAACTGTTT 600
Db      541  CTGTTACATCATCGACATTAAACCGGACCTTTATTGAGATAGGCGCGTGTGCTTAACAGAACTGTTT 600
Qy      601  TGGGGCGACGTTCCCGCTCGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTT 660
Db      601  TGGGGCGACGTTCCCGCTCGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTT 660
Qy      661  TTACCGGCGTCACCGTTGTACTAA 684
Db      661  TTACCGGCGTCACCGTTGTACTAA 684
```

```
RESULT 3
AX329368  AX329368  684 bp  DNA  linear  PAT 08-JAN-2002
LOCUS     Sequence 7 from Patent WO0194607.
DEFINITION
ACCESSION AX329368
VERSION   AX329368.1  GI:18102367
KEYWORDS  'synthetic construct
          'synthetic construct
          other sequences; artificial sequences.'
SOURCE    synthetic construct
ORGANISM  synthetic construct
REFERENCE 1
AUTHORS  Meyer, K., van Dyk, D.E. and Viitanen, P.V.
TITLE    High level production of p-hydroxybenzoic acid in green plants
JOURNAL  Patent: WO 0194607-A 7 13-DEC-2001;
        E.I. Dupont De Nemours (US)
FEATURES  Location/Qualifiers
           source
             1..684
             /organism="synthetic construct"
             /mol_type="unassigned DNA"
             /db_xref="taxon:32630"
             /note="synthetic CPL"
```

```
ORIGIN
Query Match      100.0%; Score 684; DB 2; Length 684;
Best Local Similarity 100.0%; Pred. No. 6.8e-202;
Matches 684; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  ATGGCTTCTCTGTGCTATTCTTCAGCAGCTGTTGCCACAGCAGCAATGTTACACAAGCT 60
Db      1  ATGGCTTCTCTGTGCTATTCTTCAGCAGCTGTTGCCACAGCAGCAATGTTACACAAGCT 60
Qy      61  AGCATGGTTGCAACCTTTCACCTGGTCTCAAAATCTTCAGGCACCTTTCCCTGTTTACAAAG 120
Db      61  AGCATGGTTGCAACCTTTCACCTGGTCTCAAAATCTTCAGGCACCTTTCCCTGTTTACAAAG 120
Qy      121  CAAAACCTTGACATCATCTTCATGCTAGCAATGGTGAAGAGTTAGCTGATGAGGTG 180
```

Db 121 CAAAACCTTGACATCACTCCATTTGCTAGCAATGGTGGAGAGTTAGCTGCATGCAAGTG 180  
Qy 181 TGGCATATGTCAACACCCCGCTTAACGCAACTGCGTGGCTGGCTATTGTTAAAGAGATC 240  
Db 181 TGGCATATGTCAACACCCCGCTTAACGCAACTGCGTGGCTGGCTATTGTTAAAGAGATC 240  
Qy 241 CTGGCCCTGGATCCGCAACTGCTGCACTGGCTGGCTGGATGATGACAAACGTT 300  
Db 241 CTGGCCCTGGATCCGCAACTGCTGCACTGGCTGGCTGGATGATGACAAACGTT 300  
Qy 301 TTTGACACAGGGAAGAAAGCTGAGCGTGAAGCTGACGATGATCCGGAAGGGTTTGTGAGCAG 360  
Db 301 TTTGACACAGGGAAGAAAGCTGAGCGTGAAGCTGACGATGATCCGGAAGGGTTTGTGAGCAG 360  
Qy 361 AATGAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTTACGTGAA 420  
Db 361 AATGAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTTACGTGAA 420  
Qy 421 ATTTTGTATGTGCGGATGGTGAA CCGTGGCTTGGCGGTGCGTACCGTCTGCTGTGTCA 480  
Db 421 ATTTTGTATGTGCGGATGGTGAA CCGTGGCTTGGCGGTGCGTACCGTCTGCTGTGTCA 480  
Qy 481 AGTTAAGCGGCGCGGAGCTGGCGTTACAAAATTGGGTAAACCGCGTTAGGACGCTAT 540  
Db 481 AGTTAAGCGGCGCGGAGCTGGCGTTACAAAATTGGGTAAACCGCGTTAGGACGCTAT 540  
Qy 541 CTGTTTCAATCATCGACATTAACCCGGGACTTTATTAGATAGGCGCGTATGCCGGCTG 600  
Db 541 CTGTTTCAATCATCGACATTAACCCGGGACTTTATTAGATAGGCGCGTATGCCGGCTG 600  
Qy 601 TGGGGCGACGTTCCCGCTCGGATTAACCGGTAAACCGCTGTTGCTAAACAGAACTGTTT 660  
Db 601 TGGGGCGACGTTCCCGCTCGGATTAACCGGTAAACCGCTGTTGCTAAACAGAACTGTTT 660  
Qy 661 TTACCGGCGTCAACCGTTGACTAA 684  
Db 661 TTACCGGCGTCAACCGTTGACTAA 684

RESULT 4  
AX382258 AX382258 684 bp DNA linear PAT 18-MAR-2002  
LOCUS Sequence 16 from Patent WO0204653.  
DEFINITION  
ACCESSION AX382258  
VERSION AX382258.1 GI:19577037  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1  
AUTHORS Flint, D., Meyer, K. and Viitanen, P.V.  
TITLE Sinapolyglucose:maltate sinapolytransferase form maltate conjugates from benzoic acid glucosides  
JOURNAL Patent: WO 0204653-A 16 17-JAN-2002;  
FEATURES E.I. DUPONT DE NEUMOURS AND COMPANY (US)  
Location/Qualifiers  
source  
1. .684  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="open reading frame of the chloroplast-targeted CPL fusion protein"

ORIGIN  
Query Match 100.0%; Score 684; DB 2; Length 684;  
Best Local Similarity 100.0%; Pred. No. 6.8e-202;  
Matches 684; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ATGGGCTCTCTGTGATTTCTTTTTCAGCAGCTGTTGCCACACGAGCAATGTTACACAAGCT 60  
Db 1 ATGGGCTCTCTGTGATTTCTTTTTCAGCAGCTGTTGCCACACGAGCAATGTTACACAAGCT 60  
Qy 61 AGCATGGTGGACCTTTTCACTGGTCTCAAATCTTCAGCCACTTTCCCTGTTTACAAAGAAG 120

Db 61 AGCATGGTGGACCTTTCACTGGTCTCAAATCTTCAGCCACTTTCCCTGTTTACAAAGAAG 120  
Qy 121 CAAAACCTTGACATCACTCCATTTGCTAGCAATGGTGGAGAGTTAGCTGCATGCAAGTG 180  
Db 121 CAAAACCTTGACATCACTCCATTTGCTAGCAATGGTGGAGAGTTAGCTGCATGCAAGTG 180  
Qy 181 TGGCATATGTCAACACCCCGCTTAACGCAACTGCGTGGCTGGCTATTGTTAAAGAGATC 240  
Db 181 TGGCATATGTCAACACCCCGCTTAACGCAACTGCGTGGCTGGCTATTGTTAAAGAGATC 240  
Qy 241 CTGGCCCTGGATCCGCAACTGCTGCACTGGCTGGCTGGATGATGACAAACGTT 300  
Db 241 CTGGCCCTGGATCCGCAACTGCTGCACTGGCTGGCTGGATGATGACAAACGTT 300  
Qy 301 TTTGACACAGGGAAGAAAGCTGAGCGTGAAGCTGACGATGATCCGGAAGGGTTTGTGAGCAG 360  
Db 301 TTTGACACAGGGAAGAAAGCTGAGCGTGAAGCTGACGATGATCCGGAAGGGTTTGTGAGCAG 360  
Qy 361 AATGAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTTACGTGAA 420  
Db 361 AATGAAATCCCGAAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTTACGTGAA 420  
Qy 421 ATTTTGTATGTGCGGATGGTGAA CCGTGGCTTGGCGGTGCGTACCGTCTGCTGTGTCA 480  
Db 421 ATTTTGTATGTGCGGATGGTGAA CCGTGGCTTGGCGGTGCGTACCGTCTGCTGTGTCA 480  
Qy 481 AGTTAAGCGGCGCGGAGCTGGCGTTACAAAATTGGGTAAACCGCGTTAGGACGCTAT 540  
Db 481 AGTTAAGCGGCGCGGAGCTGGCGTTACAAAATTGGGTAAACCGCGTTAGGACGCTAT 540  
Qy 541 CTGTTTCAATCATCGACATTAACCCGGGACTTTATTAGATAGGCGCGTATGCCGGCTG 600  
Db 541 CTGTTTCAATCATCGACATTAACCCGGGACTTTATTAGATAGGCGCGTATGCCGGCTG 600  
Qy 601 TGGGGCGACGTTCCCGCTCGGATTAACCGGTAAACCGCTGTTGCTAAACAGAACTGTTT 660  
Db 601 TGGGGCGACGTTCCCGCTCGGATTAACCGGTAAACCGCTGTTGCTAAACAGAACTGTTT 660  
Qy 661 TTACCGGCGTCAACCGTTGACTAA 684  
Db 661 TTACCGGCGTCAACCGTTGACTAA 684

RESULT 5  
AX382258 AX382258 854 bp DNA linear PAT 08-AUG-2001  
LOCUS Sequence 24 from patent US 6210937.  
DEFINITION  
ACCESSION AX382258  
VERSION AX382258.1 GI:15106655  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 854)  
AUTHORS Ward, T.E., Watkins, C.S., Bulmer, D.K., Johnson, B.F. and Amaratunga, M.  
TITLE Development of genetically engineered bacteria for production of selected aromatic compounds  
JOURNAL Patent: US 6210937-A 24 03-APR-2001;  
FEATURES Location/Qualifiers  
source  
1. .854  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN  
Query Match 72.8%; Score 498; DB 2; Length 854;  
Best Local Similarity 100.0%; Pred. No. 9.5e-144;  
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 187 ATGTCACACCCCGCTTAACGCAACTGCGTGGCTGCGCTATTGTTAAAGAGATCCCTGCC 246  
Db 321 ATGTCACACCCCGCTTAACGCAACTGCGTGGCTGCGCTATTGTTAAAGAGATCCCTGCC 380



```
RESULT 7
LOCUS      E11274
DEFINITION ubiC and ubiA gene.
ACCESSION E11274
VERSION   E11274.1 GI:22024916
KEYWORDS  JP 1996107789-A/1.
SOURCE    Escherichia coli
ORGANISM  Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
           Enterobacteriaceae; Escherichia.
REFERENCE Matsuda.H., Kawamuki.M. and Nakagawa.T.
AUTHORS   PRODUCTION OF UBIQUINONE-10
TITLE     Patent: JP 1996107789-A 1 30-APR-1996;
JOURNAL   ALPHA- SHOKUHIN KK
COMMENT   OS Escherichia coli
          FN JP 1996107789-A/1
          PD 30-APR-1996
          PF 13-OCT-1994 JP 1994273071
          PI MATSUDA HIDEYUKI, KAWAMUKI MAKOTO, NAKAGAWA TSUYOSHI PC
          CL2N15/09,A61K31/12,A61K31/12,C07H21/04,Cl2N1/21,Cl2P7/66, PC
          (Cl2N1/21,
          PC Cl2R1:01), (Cl2P7/66,Cl2R1:01);
          CC strandedness: Double;
          CC topology: Linear;
          FH Key Location/Qualifiers
          FH source 1..2000
          FT CDS /organism='Escherichia coli'
          FT CDS 380..877
          FT CDS /product='UbiC protein'
          FT CDS 890..1762
          FT CDS /product='UbiA protein'.
          FT Location/Qualifiers
FEATURES   source
            1..2000
            /organism="Escherichia coli"
            /mol_type="genomic DNA"
            /db_xref="taxon:562"
ORIGIN
Query Match 72.8%; Score 498; DB 2; Length 2000;
Best Local Similarity 100.0%; Pred. No. 1.2e-143;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 187 ATGTACACCCCGCGTTAAGCAACTGCGTGGCTCGCTATTTGTAAGAGATCCCTGCC 246
DB 380 ATGTACACCCCGCGTTAAGCAACTGCGTGGCTCGCTATTTGTAAGAGATCCCTGCC 439
QY 247 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGA 306
DB 440 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGA 499
QY 307 CAGCAGGGAAAAACGGTAAGCGTGACGATGATCCGCGAAGGGTTTGTGAGCAGAGATGAA 366
DB 500 CAGCAGGGAAAAACGGTAAGCGTGACGATGATCCGCGAAGGGTTTGTGAGCAGAGATGAA 559
QY 367 ATCCCGGAAGAACTGCGCGTGTGCGGAAAGAGTCTGTTACTGGTTAGTGAAATTTTG 426
DB 560 ATCCCGGAAGAACTGCGCGTGTGCGGAAAGAGTCTGTTACTGGTTAGTGAAATTTTG 619
QY 427 TTATGTGCGGATGGTGAACCGTGGCTGTGCGGTGCTGACCGTGTCTCTGTGTCACGTTA 486
DB 620 TTATGTGCGGATGGTGAACCGTGGCTGTGCGGTGCTGACCGTGTCTCTGTGTCACGTTA 679
QY 487 AGCGGCGCGAGCTGCGGTTACAAAAATTTGGGTAAACCGCGTTAGGACGCTATCTGTTTC 546
DB 680 AGCGGCGCGAGCTGCGGTTACAAAAATTTGGGTAAACCGCGTTAGGACGCTATCTGTTTC 739
QY 547 ACATCATCAGACATTACCCGGGACTTTATTTAGATAGGCGGTGATGCCGGGCTGTGGGG 606
DB 740 ACATCATCAGACATTACCCGGGACTTTATTTAGATAGGCGGTGATGCCGGGCTGTGGGG 799
```

```
607 CGACGTTCCCGCTCGGATTAAAGCGGTAACCGCTGTTGCTTAACAGAACTGTTTTTACCG 666
      |||||||
800 CGACGTTCCCGCTCGGATTAAAGCGGTAACCGCTGTTGCTTAACAGAACTGTTTTTACCG 859
      |||||||

667 GCGTCACCGTTGTACTAA 684
      |||||||
860 GCGTCACCGTTGTACTAA 877
      |||||||

RESULT 8
LOCUS      ECOUBICA
DEFINITION Escherichia coli chorismate lyase (ubiC), 4-hydroxybenzoate
            octaprenyl transferase (ubiA) genes, complete cds, and
            sn-glycerol-3-phosphate acyltransferase (plsB) genes, 3' end.
ACCESSION M93136
VERSION   M93136.1 GI:148099
KEYWORDS  4-hydroxybenzoate octaprenyl transferase; chorismate lyase; plsB
            gene; sn-glycerol-3-phosphate acyltransferase; ubiA gene; ubiC
            gene.
SOURCE    Escherichia coli W3110
ORGANISM  Escherichia coli W3110
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE Lightner,V.A., Bell,R.M. and Modrich,P.
AUTHORS   The DNA sequences encoding plsB and dgk loci of Escherichia coli
JOURNAL   J. Biol. Chem. 258 (18), 10856-10861 (1983)
COMMENT   2 (bases 1 to 2000)
            Nichols,B.P. and Green,J.M.
            Cloning and Sequencing of Escherichia coli ubiC and purification of
            chorismate lyase
            Unpublished (1992)
FEATURES   source
            1..2000
            /organism="Escherichia coli W3110"
            /mol_type="genomic DNA"
            /strain="K-12"
            /sub_strain="W3110"
            /db_xref="taxon:316407"
            /map="91.8 min"
            /tissue.lib="Kohare lambda-1F8"
            380..877
            /gene="ubiC"
            380..877
            /gene="ubiC"
            /function="enzymatic chorismate --> p-hydroxybenzoate +
            pyruvate"
            /experiment="experimental evidence, no additional details
            recorded"
            /citation=[2]
            /codon_start=1
            /transl_table=11
            /product="chorismate lyase"
            /protein_id="AAA24711.1"
            /db_xref="GI:148100"
            /translation="MSHPALTQLRALRYCKEIPALDQLLDWLLLEDSMTKRFEOQKG
            /VSVMIREGFVEQNEIPEELPLPKESRYLWREILLCADGEPWLAGRTVWPVSTLSG
            PELALQKLKGPLGRYLTSSLTTRDFIEIGRDAGLWGRSRRLRLSGKPLLLTELPLP
            ASPLX"
            890..1762
            /gene="ubiA"
            890..1762
            /gene="ubiA"
            /function="enzymatic - 3-octaprenyl-4-hydroxybenzoate
            synthesis"
            /experiment="experimental evidence, no additional details
            recorded"
            /citation=[2]
```

```

/codon_start=1
/transl_table=11
/product="4-hydroxybenzoate-octaprenyl transferase"
/protein_id="AAA24712.1"
/db_xref="GI:148101"
/translation="MWSLTQNKLLAFHRLMRTDKPIGALLLLWPTLWALWVATPGVP
OLWILAVFVAGVWLMRAAGCVNDYADRFKDFGHVKRTANRPLPSGAVTEKARALFVV
LVLSFLVLTNTMTILLSAALALAWYVPMKRYTHLPQVIVLGARFGHSIPNAFAA
VSEVPLSCWLMFLANIUWAVDTQYAMVDRDDDKIGIKSTAILFGQYDKLIIGIL
QIGVLAWMAIIGELNGWGYWSILVAGALFVYQOKLIANRERECFKAFMNNYVG
LVFLGLAMSYWHF"
complement(1917. .2000)
/gene="plB8"
complement(1917. .2000)
/EC_number="2.3.1.15"
/function="lysophosphatidic acid synthesis"
/experiment="experimental evidence, no additional details
recorded"
/citation=[1]
/codon_start=1
/transl_table=11
/product="en-glycerol-3-phosphate acyltransferase"
/protein_id="AAA24713.1"
/db_xref="GI:148102"
/translation="MKVYQLLAELITSDVRLTIRSATQGE"

Query Match      72.8%; Score 498; DB 15; Length 2000;
Best Local Similarity 100.0%; Pred. No. 1.2e-143;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 ATGTCACACCCCGGCTTAACGCACTGGCGCTGCGCTATTGTAAAGAGATCCCTGCC 246
DB 380 ATGTCACACCCCGGCTTAACGCACTGGCGCTGCGCTATTGTAAAGAGATCCCTGCC 439
QY 247 CTGGATCCGCAACTCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTCAA 306
DB 440 CTGGATCCGCAACTCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTCAA 499
QY 307 CAGCAGGGAAACCGGTAAAGCGTGACGATGATCCCGGAAGGTTTGTCCAGCAGAAATGAA 366
DB 500 CAGCAGGGAAACCGGTAAAGCGTGACGATGATCCCGGAAGGTTTGTCCAGCAGAAATGAA 559
QY 367 ATCCCGGAAGAACTCCCGCTGCTCGCGAAAGAGTCTCGTTACTGTTACGTGAATTTTG 426
DB 560 ATCCCGGAAGAACTCCCGCTGCTCGCGAAAGAGTCTCGTTACTGTTACGTGAATTTTG 619
QY 427 TTATGTCCGATGGTGAACCGTGGCTTCCCGTGTACCGTCTGTTCTGTGTCAACGTTA 486
DB 620 TTATGTCCGATGGTGAACCGTGGCTTCCCGTGTACCGTCTGTTCTGTGTCAACGTTA 679
QY 487 AGCGGCGCGGAGCTGGCGTTACAAAATTTGGGTAAACCGCGTTAGGACGCTATCTGTTTC 546
DB 680 AGCGGCGCGGAGCTGGCGTTACAAAATTTGGGTAAACCGCGTTAGGACGCTATCTGTTTC 739
QY 547 ACATCATCGACATTAACCGGACCTTTATAGATAGGCGCGGTGATGCCGGGCTGTGGGG 606
DB 740 ACATCATCGACATTAACCGGACCTTTATAGATAGGCGCGGTGATGCCGGGCTGTGGGG 799
QY 607 CGACGTTCCCGCTCGGCTAAGCGGTAAACCGGTTTGCTAAACAGAACTGTTTTACCG 666
DB 800 CGACGTTCCCGCTCGGCTAAGCGGTAAACCGGTTTGCTAAGCGAACTGTTTTACCG 859
QY 667 GCGTCACCGTTGTACTAA 684
DB 860 GCGTCACCGTTGTACTAA 877
```

RESULT 9  
ECOUBIPLS  
LOCUS Escherichia coli 2000 bp DNA linear BCT 26-APR-1993  
DEFINITION Escherichia coli 4-hydroxybenzoate octaprenyl transferase (ubiA)

```

ACCESSION M93413.1 GI:148106
VERSION M93413.1
KEYWORDS 4-hydroxybenzoate octaprenyl transferase; chorismate lyase; plB8
gene; en-glycerol-3-phosphate acyltransferase; ubiA gene; ubiC
gene;
Escherichia coli W3110
Escherichia coli W3110
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
1 (bases 1 to 2000)
Lightner,V.A., Bell,R.M. and Modrich,P.
The DNA sequences encoding plB8 and dgk loci of Escherichia coli
J. Biol. Chem. 258 (18), 10856-10861 (1983)
6309817
2 (bases 1 to 2000)
Nichols,B.P. and Green,J.M.
Cloning and sequencing of Escherichia coli ubiC and purification of
Chorismate lyase
J. Bacteriol. 174 (16), 5309-5316 (1992)
1644758
COMMENT Original source text: Escherichia coli (sub_strain W3110, strain
K-12) (library: Kohara lambda) DNA.
FEATURES
Location/Qualifiers
source
1. .2000
/organism="Escherichia coli W3110"
/mol_type="genomic DNA"
/strain="K-12"
/sub_strain="W3110"
/db_xref="taxon:316407"
/map="91.8 min"
/tissue_lib="Kohara lambda"
380. .877
/gene="ubiC"
380. .877
/gene="ubiC"
/function="enzymatic chorismate --> p-hydroxybenzoate +
pyruvate"
/experiment="experimental evidence, no additional details
recorded"
/notes="N-terminal amino acid sequence of chorismate lyase
:NH2-SHPALTQLRALRYCEIPAL-"
/citation=[2]
/codon_start=1
/transl_table=11
/product="chorismate lyase"
/protein_id="AAA24716.1"
/db_xref="GI:148107"
/translation="MSHPALTOQLRALRYCKEIPALDPQLLDWLLLEDSTMTRFBOQKG
TVSVTMIREGFVEQNEIPEELPLLPKESRYWLREILLCADGEPWLAGRTVYVPVSTLSG
PEALQKLKGLTGLGRYLTSTLTDRDFIEGRDAGLWRRRLRSLGKPLLTLEFLP
ASPLY"
890. .1762
/gene="ubiA"
890. .1762
/gene="ubiA"
/function="enzymatic - 3-octaprenyl-4-hydroxybenzoate
synthesis"
/experiment="experimental evidence, no additional details
recorded"
/citation=[2]
/number=2
/codon_start=1
/transl_table=11
/product="4-hydroxybenzoate-octaprenyl transferase"
/protein_id="AAA24717.1"
/db_xref="GI:148108"
/translation="MWSLTQNKLLAFHRLMRTDKPIGALLLLWPTLWALWVATPGVP
OLWILAVFVAGVWLMRAAGCVNDYADRFKDFGHVKRTANRPLPSGAVTEKARALFVV
LVLSFLVLTNTMTILLSAALALAWYVPMKRYTHLPQVIVLGARFGHSIPNAFAA
VSEVPLSCWLMFLANIUWAVDTQYAMVDRDDDKIGIKSTAILFGQYDKLIIGIL
QIGVLAWMAIIGELNGWGYWSILVAGALFVYQOKLIANRERECFKAFMNNYVG
```

gene  
CDS  
LVLFLGLAMSYWHP"  
complement (1917. .2000)  
/gene="pleB"  
complement (1917. .2000)  
/gene="pleB"  
/EC\_number="2.3.1.15"  
/function="lysophosphatidic acid synthesis"  
/experiment="experimental evidence, no additional details  
recorded"  
/citation=[1]  
/codon\_start=1  
/transl\_table=11  
/product="sn-glycerol-3-phosphate acyltransferase"  
/protein\_id="AAA24718.1"  
/db\_xref="GI:148109"  
/translation="MKVYQLLAELITSDVRLTIESATQSEG"

ORIGIN

Query Match 72.8%; Score 498; DB 15; Length 2000;  
Best Local Similarity 100.0%; Pred. No. 1.2e-143;  
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 ATGTACACCCCGCTTAACGCACTGCGTGGCTGGCTATTGTAAGAGATCCCTGCC 246  
DB 380 ATGTACACCCCGCTTAACGCACTGCGTGGCTGGCTATTGTAAGAGATCCCTGCC 439  
QY 247 CTGGATCCGCACTGCTGCACTGGCTGGCTGGAGATTCCATGACAAACGTTTGA 306  
DB 440 CTGGATCCGCACTGCTGCACTGGCTGGCTGGAGATTCCATGACAAACGTTTGA 499  
QY 307 CAGCAGGGAAAAACGTAAGCGTGACGATATCCGCGAAGGGTTTCTCGAGCAGAAATGA 366  
DB 500 CAGCAGGGAAAAACGTAAGCGTGACGATATCCGCGAAGGGTTTCTCGAGCAGAAATGA 559  
QY 367 ATCCCGGAGAACTGCGCTGCTGCGAAGAGTCTGTTACTGTTAGTGAATTTTG 426  
DB 560 ATCCCGGAGAACTGCGCTGCTGCGAAGAGTCTGTTACTGTTAGTGAATTTTG 619  
QY 427 TTATGTGCGGATGGTGAACCGTGGCTTGGCGGTCGTACCGTCTGTTCTGTCAACGTTA 486  
DB 620 TTATGTGCGGATGGTGAACCGTGGCTTGGCGGTCGTACCGTCTGTTCTGTCAACGTTA 679  
QY 487 AGCGGCGCGAGCTGCGTTACAAAAATTGGGTAAACCCCGTTAGGACGCTATCTGTT 546  
DB 680 AGCGGCGCGAGCTGCGTTACAAAAATTGGGTAAACCCCGTTAGGACGCTATCTGTT 739  
QY 547 ACATCATGCACATTACCCGGGACTTTATGAGATAGGCGGTGATCCCGGCTGTGGGG 606  
DB 740 ACATCATGCACATTACCCGGGACTTTATGAGATAGGCGGTGATCCCGGCTGTGGGG 799  
QY 607 CGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTACCG 666  
DB 800 CGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTACCG 859  
QY 667 GCGTACCGCTGTACTAA 684  
DB 860 GCGTACCGCTGTACTAA 877

RESULT 10  
ECUBIAC  
LOCUS  
DEFINITION  
E.coli ubiC and ubiA genes for chorismate lyase and  
4-hydroxybenzoate octaprenyltransferase.  
ACCESSION  
X57434  
VERSION  
X57434.1 GI:43233  
KEYWORDS  
4-hydroxybenzoate octaprenyl transferase; 4-hydroxybenzoate  
synthetase; chorismate lyase.  
SOURCE  
Escherichia coli  
ORGANISM  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
REFERENCE  
1

AUTHORS  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
Nishimura, K., Nakahigashi, K. and Inokuchi, H.  
Location of the ubiA gene on the physical map of Escherichia coli  
J. Bacteriol. 174 (17), 5762 (1992)  
1512213  
2 (bases 1 to 2034)  
Nishimura, K.  
Direct Submission  
Submitted (28-JAN-1991) K. Nishimura, Faculty of Sciences, Kyoto  
University, Sakyo-ku, Kyoto 606, Japan  
FEATURES  
Location/Qualifiers  
1. .2034  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/strain="K12 W3110"  
/db\_xref="taxon:562"  
/map="91.7 min"  
/clone="634"  
/clone\_lib="Kohara library"  
-35\_signal  
137.142  
-10\_signal  
159.164  
gene  
291.899  
/gene="ubiC"  
291.899  
/gene="ubiC"  
/codon\_start=1  
/transl\_table=11  
/product="4-hydroxybenzoate synthetase"  
/protein\_id="CAA0681.1"  
/db\_xref="GI:43234"  
/db\_xref="GOA:P26602"  
/db\_xref="UniProtKB/Swiss-Prot:P26602"  
/translation="NRLRFCCVLDHLICTSPVNTFLRYNAFTLCNGBFGMSHPALT  
QLRLYCKEIPALDPQLLDLLEDSTKREQOGKTVSTMIRGEFVQNEIPREL  
PLPKRESYWLREILLCLADGEPWLAGRTVVPVSTLSGPELAKLQKTPGLRILFTSS  
TLTRFIEIGRDAGLWGRSLRLSGKLLELFLPASPLY"  
912.1784  
/gene="ubiA"  
912.1784  
/gene="ubiA"  
/codon\_start=1  
/transl\_table=11  
/product="4-hydroxybenzoate octaprenyltransferase"  
/protein\_id="CAA0682.1"  
/db\_xref="GI:43235"  
/db\_xref="GOA:P26601"  
/db\_xref="UniProtKB/Swiss-Prot:P26601"  
/translation="NEWSLTQNKLLAFHRLMTDRPIGALLLLWPTLWALWATPGVP  
QWLILAVFVAGVWLMRAAGCVVDYADRFKDHVXRTANRPLPSGAVTKEKALFV  
LVLIISPLVLTNTMTILLSTAAALAWYPPMKRYTHLPQVVLGAAFGWSTPMAFAA  
VSESVPKSCWLMFLANILWAVDTQYAMVDRDDVKIGIKSTAILFGQYDKLIIGIL  
QIGVLALMAIIGELNGLGWYYSILVAGALFVYQKLIANRERBACPKAFMNNVVG  
LVLFLGLAMSYWHP"

ORIGIN

Query Match 72.8%; Score 498; DB 15; Length 2034;  
Best Local Similarity 100.0%; Pred. No. 1.2e-143;  
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 ATGTACACCCCGCTTAACGCACTGCGTGGCTGGCTATTGTAAGAGATCCCTGCC 246  
DB 402 ATGTACACCCCGCTTAACGCACTGCGTGGCTGGCTATTGTAAGAGATCCCTGCC 461  
QY 247 CTGGATCCGCACTGCTGCACTGGCTGGCTGGAGATTCCATGACAAACGTTTGA 306  
DB 462 CTGGATCCGCACTGCTGCACTGGCTGGCTGGAGATTCCATGACAAACGTTTGA 521  
QY 307 CAGCAGGGAAAAACGTAAGCGTGACGATGATCCGCGAAGGGTTTCTCGAGCAGAAATGA 366  
DB 522 CAGCAGGGAAAAACGTAAGCGTGACGATGATCCGCGAAGGGTTTCTCGAGCAGAAATGA 581  
QY 367 ATCCCGGAGAACTGCGCTGCTGCGAAGAGTCTGTTACTGTTAGTGAATTTTG 426  
DB 582 ATCCCGGAGAACTGCGCTGCTGCGAAGAGTCTGTTACTGTTAGTGAATTTTG 641



QY 427 TTATGTCCGATGTTAAACCGTGGCTTCCCGGCTGTACCGTGTCTCTGTGTCAACGTTA 486  
Db 642 TTATGTCCGATGTTAAACCGTGGCTTCCCGGCTGTACCGTGTCTCTGTGTCAACGTTA 701  
QY 487 AGCGGGCGGAGCTCGGCTTACAAAATTGGGTAAACCGCGTTAGGACGCTATCTGTTTC 546  
Db 702 AGCGGGCGGAGCTCGGCTTACAAAATTGGGTAAACCGCGTTAGGACGCTATCTGTTTC 761  
QY 547 ACATCATCGACATTAAACCGGAGCTTTATTGAGATAGCGCGTGTAGTCCGGGCTGTGGGG 606  
Db 762 ACATCATCGACATTAAACCGGAGCTTTATTGAGATAGCGCGTGTAGTCCGGGCTGTGGGG 821  
QY 607 CGAGTTCCTCCGCTCGGATTAAGCGGTAAACCGCTGTGCTTAAACAGAACTGTTTTACCG 666  
Db 822 CGAGTTCCTCCGCTCGGATTAAGCGGTAAACCGCTGTGCTTAAACAGAACTGTTTTACCG 881  
QY 667 GCGTCACCGTGTACTAA 684  
Db 882 GCGTCACCGTGTACTAA 899

RESULT 11  
ECUBI  
LOCUS 2348 bp DNA linear BCT 09-SEP-2004  
DEFINITION E.coli genes ubiC and ubiA.  
ACCESSION X66619.1  
VERSION X66619.1 GI:43230  
KEYWORDS 4-hydroxybenzoate-octaprenyltransferase; chorismate-pyruvate lyase; ubiA gene; ubiC gene.  
SOURCE Escherichia coli  
ORGANISM Escherichia coli  
REFERENCE 1 (bases 1 to 2348)  
AUTHORS Siebert, M., Rechold, A., Melzer, M., May, U., Berger, U., Schroder, G., Schroder, J., Severin, K. and Heide, L.  
TITLE Ubiquinone biosynthesis. Cloning of the genes coding for chorismate pyruvate-lyase and 4-hydroxybenzoate octaprenyl transferase from Escherichia coli  
JOURNAL FEBS Lett. 307 (3), 347-350 (1992)  
PUBMED 1644192  
REFERENCE 2 (bases 1 to 2348)  
AUTHORS Heide, L.  
TITLE Direct Submission  
JOURNAL Submitted (04-JUN-1992) L. Heide, Inst of Pharmaceutical Biology, Schaenzlestrasse 1, 7800 Freiburg, FRG  
COMMENT On Sep 24, 2004 this sequence version replaced gi:2531104.  
See also K00127, M93136 & M93413.  
FEATURES  
source  
1..2348  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/strain="MC4100"  
/db\_xref="taxon:562"  
/map="91.8 mins"  
/clone="pALMUI"  
/clone\_lib="Sau3A partial digested DNA in pBluescript"  
558..563  
750..1067  
/gene="ubiC"  
750..1067  
/gene="ubiC"  
/codon\_start=1  
/product="chorismate-pyruvate lyase"  
/protein\_id="CAA47181.1"  
/db\_xref="GI:43231"  
/db\_xref="GOA:P26602"  
/translation="MSHPALTQLRALRYCKEIPALDPQLLDLWLLIEDSMTRKRFQOQK TVSVTMIREGFVEQNEIPEELPLLPKESRYWLREILLCADGEPWLAGRTVVPVSTLSG PELALQKLGKTPLGRLYFTSITLDRDFEIGRDAGLWGRRLSLRSLGKPLLLTELPLP

RBS  
gene  
CDS

ASPLV"  
1068..1073  
1080..1952  
/gene="ubiA"  
1080..1952  
/gene="ubiA"  
/codon\_start=1  
/transl\_table=11  
/product="4-hydroxybenzoate-octaprenyl transferase"  
/protein\_id="CAA47182.1"  
/db\_xref="GI:43232"  
/db\_xref="GOA:P26601"  
/db\_xref="UniProtKB/Swiss-Prot:P26601"  
/translation="MWSLTONKLLAFHRLNRTDKPIGALLLLHPTLWALWVATPGVP QWTLAVFVAGVWLMRAAGCVVNDYADKKFDGHVKRYTHLPQVVGAAFGWISIPMAFAA LVLISFLVLTANTMTITLLSIAALAAWVYPMKRYTHLPQVVGAAFGWISIPMAFAA VSEVPLSCWLMFLANILWAVDYQYAMVDRDDVKIGIKSTAILFQYDKLIIGIL QIGVLAALMAIIGELNGLWGYWSILVAGALFVYQOKLIANRERACPKAFMNNYVG LVLFLGLMSYWHF"  
1254..1322  
/gene="ubiA"  
/bound\_moiety="polyprenyl"  
1965..1989  
terminator  
repeat\_region 1994..2096  
/note="palindrome"

ORIGIN  
Query Match 72.8%; Score 498; DB 15; Length 2348;  
Best Local Similarity 100.0%; Pred. No. 1.2e-143;  
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 187 ATGTCACACCCCGGTTAAACCGCAACTGCGTGGCTGGCTATTGTAAGAGATCCCTGCC 246  
Db 570 ATGTCACACCCCGGTTAAACCGCAACTGCGTGGCTGGCTATTGTAAGAGATCCCTGCC 629  
QY 247 CTGGATCCGAACCTGCTGACTGGCTGTTGCTGGAGGATTCATGACAAAAAGTTTGA 306  
Db 630 CTGGATCCGAACCTGCTGACTGGCTGTTGCTGGAGGATTCATGACAAAAAGTTTGA 689  
QY 307 CAGCAGGAAAAACCGTAAAGCGTCACGATATCCCGAAGGGTTTGTGAGCAGAAATGA 366  
Db 690 CAGCAGGAAAAACCGTAAAGCGTCACGATATCCCGAAGGGTTTGTGAGCAGAAATGA 749  
QY 367 ATCCCGAAGAACTGCCGCTGCTCCGAAAGAGTCTGTTACTGTTACGTGAATTTTG 426  
Db 750 ATCCCGAAGAACTGCCGCTGCTCCGAAAGAGTCTGTTACTGTTACGTGAATTTTG 809  
QY 427 TTATGTCCGATGTTAAACCGTGGCTTCCCGGCTGTACCGTGTCTCTGTGTCAACGTTA 486  
Db 810 TTATGTCCGATGTTAAACCGTGGCTTCCCGGCTGTACCGTGTCTCTGTGTCAACGTTA 869  
QY 487 AGCGGGCGGAGCTGGGCTTACAAAATTGGGTAAACCGCGTTAGGACGCTATCTGTTTC 546  
Db 870 AGCGGGCGGAGCTGGGCTTACAAAATTGGGTAAACCGCGTTAGGACGCTATCTGTTTC 929  
QY 547 ACATCATCGACATTAAACCGGAGCTTTATTGAGATAGCGCGTGTAGGACGCTGTGGGG 606  
Db 930 ACATCATCGACATTAAACCGGAGCTTTATTGAGATAGCGCGTGTAGGACGCTGTGGGG 989  
QY 607 CGAGTTCCTCCGCTCGGATTAAGCGGTAAACCGCTGTGCTTAAACAGAACTGTTTTACCG 666  
Db 990 CGAGTTCCTCCGCTCGGATTAAGCGGTAAACCGCTGTGCTTAAACAGAACTGTTTTACCG 1049  
QY 667 GCGTCACCGTGTACTAA 684  
Db 1050 GCGTCACCGTGTACTAA 1067

RESULT 12  
AR144789  
LOCUS  
DEFINITION Sequence 25 from patent US 6210937.  
ACCESSION AR144789

AR144789 6641 bp DNA linear PAT 08-AUG-2001



Query Match	Best Local Similarity	Matches	Score	DB	Length
U00096_10	1000001	1110000	72.88	4200001	1100000
U00096_11	1100001	1210000	100.0%	4200001	1100000
U00096_12	1200001	1310000	0	4200001	1100000
U00096_13	1300001	1410000	0	4200001	1100000
U00096_14	1400001	1510000	0	4200001	1100000
U00096_15	1500001	1610000	0	4200001	1100000
U00096_16	1600001	1710000	0	4200001	1100000
U00096_17	1700001	1810000	0	4200001	1100000
U00096_18	1800001	1910000	0	4200001	1100000
U00096_19	1900001	2010000	0	4200001	1100000
U00096_20	2000001	2110000	0	4200001	1100000
U00096_21	2100001	2210000	0	4200001	1100000
U00096_22	2200001	2310000	0	4200001	1100000
U00096_23	2300001	2410000	0	4200001	1100000
U00096_24	2400001	2510000	0	4200001	1100000
U00096_25	2500001	2610000	0	4200001	1100000
U00096_26	2600001	2710000	0	4200001	1100000
U00096_27	2700001	2810000	0	4200001	1100000
U00096_28	2800001	2910000	0	4200001	1100000
U00096_29	2900001	3010000	0	4200001	1100000
U00096_30	3000001	3110000	0	4200001	1100000
U00096_31	3100001	3210000	0	4200001	1100000
U00096_32	3200001	3310000	0	4200001	1100000
U00096_33	3300001	3410000	0	4200001	1100000
U00096_34	3400001	3510000	0	4200001	1100000
U00096_35	3500001	3610000	0	4200001	1100000
U00096_36	3600001	3710000	0	4200001	1100000
U00096_37	3700001	3810000	0	4200001	1100000
U00096_38	3800001	3910000	0	4200001	1100000
U00096_39	3900001	4010000	0	4200001	1100000
U00096_40	4000001	4110000	0	4200001	1100000
U00096_41	4100001	4210000	0	4200001	1100000
U00096_42	4200001	4310000	0	4200001	1100000
U00096_43	4300001	4410000	0	4200001	1100000
U00096_44	4400001	4510000	0	4200001	1100000
U00096_45	4500001	4610000	0	4200001	1100000
U00096_46	4600001	4639675	0	4200001	1100000
Continuation (43 of 47) of U00096 from base 4200001 (U00096 Escherichia coli K-12 MG165					
Query Match	Best Local Similarity	Matches	Score	DB	Length
187	ATGTCACACCCCGGTTAAAGCAACTCGTGCCTGCGCTATTGTAAGAGATCCCTGCC	246	72.88	4200001	1100000
50529	ATGTCACACCCCGGTTAAAGCAACTCGTGCCTGCGCTATTGTAAGAGATCCCTGCC	246	100.0%	4200001	1100000
247	CTGGATCCGCAACTGCTCGACTGCTGTTGCTGAGGATTCCTATGACAAACGTTTGA	306	0	4200001	1100000
50589	CTGGATCCGCAACTGCTCGACTGCTGTTGCTGAGGATTCCTATGACAAACGTTTGA	306	0	4200001	1100000
307	CAGCAGGAAACGTTAAGCGTGACCATGATCCGCAAGGTTTGTGACGACAAATGA	366	0	4200001	1100000
50649	CAGCAGGAAACGTTAAGCGTGACCATGATCCGCAAGGTTTGTGACGACAAATGA	366	0	4200001	1100000
367	ATCCCCGAACCTGCCGCTGCTGCCGCTGCTGCTTACTGTTACGTGAAATTTG	426	0	4200001	1100000
50709	ATCCCCGAACCTGCCGCTGCTGCCGCTGCTGCTTACTGTTACGTGAAATTTG	426	0	4200001	1100000
427	TTATGTCGCGATGTTGAAACCGTGCGTTCGCGTTCGTAACCGTTCCTGTCACCGTTA	486	0	4200001	110

Db 50949 CGACGTTCCCGCTCGGATTAAAGCGGTAAACCGCTGTTGCTTAACAGAACTGTTTTTACCG 51008

Qy 667 GCGTCACCGTTGTACTAA 684

Db 51009 GCGTCACCGTTGTACTAA 51026

## RESULT 14

AP009048\_42

WPCOMMENT

Sequence split into 47 fragments LOCUS AP009048 Accession AP009048

Fragment Name	Begin	End
AP009048_00	1	110000
AP009048_01	100001	210000
AP009048_02	200001	310000
AP009048_03	300001	410000
AP009048_04	400001	510000
AP009048_05	500001	610000
AP009048_06	600001	710000
AP009048_07	700001	810000
AP009048_08	800001	910000
AP009048_09	900001	1010000
AP009048_10	1000001	1110000
AP009048_11	1100001	1210000
AP009048_12	1200001	1310000
AP009048_13	1300001	1410000
AP009048_14	1400001	1510000
AP009048_15	1500001	1610000
AP009048_16	1600001	1710000
AP009048_17	1700001	1810000
AP009048_18	1800001	1910000
AP009048_19	1900001	2010000
AP009048_20	2000001	2110000
AP009048_21	2100001	2210000
AP009048_22	2200001	2310000
AP009048_23	2300001	2410000
AP009048_24	2400001	2510000
AP009048_25	2500001	2610000
AP009048_26	2600001	2710000
AP009048_27	2700001	2810000
AP009048_28	2800001	2910000
AP009048_29	2900001	3010000
AP009048_30	3000001	3110000
AP009048_31	3100001	3210000
AP009048_32	3200001	3310000
AP009048_33	3300001	3410000
AP009048_34	3400001	3510000
AP009048_35	3500001	3610000
AP009048_36	3600001	3710000
AP009048_37	3700001	3810000
AP009048_38	3800001	3910000
AP009048_39	3900001	4010000
AP009048_40	4000001	4110000
AP009048_41	4100001	4210000
AP009048_42	4200001	4310000
AP009048_43	4300001	4410000
AP009048_44	4400001	4510000
AP009048_45	4500001	4610000
AP009048_46	4600001	4646332

Continuation (43 of 47) of AP009048 from base 4200001 (AP009048 Escherichia coli W3110 D

Query Match 72.8%; Score 498; DB 15; Length 110000;

Best Local Similarity 100.0%; Pred. No. 3e-143; Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 187 ATGTACACACCCCGCTTAACGCAACTGGCTGGCTGCGCTATTGTAAAGAGATCCCTGCC 246

Db 56096 ATGTACACACCCCGCTTAACGCAACTGGCTGGCTGCGCTATTGTAAAGAGATCCCTGCC 56155

Qy 247 CTGGATCCGCAACTGCTGCACTGGCTGTGTGTGGAGAGATTCATGACAAAACGTTTTGAA 306

Db 56156 CTGGATCCGCAACTGCTGCACTGGCTGTGTGTGGAGAGATTCATGACAAAACGTTTTGAA 56215

Qy	307	CAGCAGGGAAGAAACGGTAAAGCGTGAAGCGATATCCGGAAGGGTTTGTTCGACGAGATGAA	366
Db	56216	CAGCAGGGAAGAAACGGTAAAGCGTGAAGCGTGAAGCGTAAAGGGTTTGTTCGACGAGATGAA	56275
Qy	367	ATCCCGGAAGAACTGCCGCTGCTGCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTTG	426
Db	56276	ATCCCGGAAGAACTGCCGCTGCTGCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTTG	56335
Qy	427	TTATGTGCGCATGGTGAAACCGTGGCTTGCCTCCGCTCGTACCGTCTGTTGTCACAGTTA	486
Db	56336	TTATGTGCGCATGGTGAAACCGTGGCTTGCCTCCGCTCGTACCGTCTGTTGTCACAGTTA	56395
Qy	487	AGCGGCGGAGCTGGCGTTTACAAAAATTTGGTAAACCGCTTAGGACGCTATCTGTTTC	546
Db	56396	AGCGGCGGAGCTGGCGTTTACAAAAATTTGGTAAACCGCTTAGGACGCTATCTGTTTC	56455
Qy	547	ACATCATCGACATTAACCCCGGACTTTTATTGAGATAGGCCGCTGATGCCGGCTGTGGGG	606
Db	56456	ACATCATCGACATTAACCCCGGACTTTTATTGAGATAGGCCGCTGATGCCGGCTGTGGGG	56515
Qy	607	CGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTACCG	666
Db	56516	CGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTACCG	56575
Qy	667	GCGTCACCGTTGTACTAA 684	
Db	56576	GCGTCACCGTTGTACTAA 56593	

## RESULT 15

ECOUW89

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

ECOUW89 176195 bp DNA linear BCT 17-DEC-1993  
E. coli chromosomal region from 89.2 to 92.8 minutes.

U00006  
U00006.1 GI:409785

Escherichia coli  
Escherichia coli

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.

1 (bases 1 to 176195)  
Blattner, F.R., Burland, V., Plunkett, G. III, Sofia, H.J. and  
Daniels, D.L.

Analysis of the Escherichia coli genome. IV. DNA sequence of the  
region from 89.2 to 92.8 minutes

Nucleic Acids Res. 21 (23), 5408-5417 (1993)  
8265357

2 (bases 1 to 176195)  
Blattner, F.R.

Direct Submission  
Submitted (03-SEP-1993) 608-263-7459

On Oct 29, 1993 this sequence version replaced gi:396288.  
This sequence was determined as part of the E. coli Genome Project  
(Frederick R. Blattner, director) at the University of

Wisconsin-Madison. Supported by award HG00301 from the NIH Human  
Genome Project. The entire sequence was independently determined

from E. coli MG1655; overlaps and conflicts with other sequence  
determinations are annotated. The start of this entry overlaps the

end of the entry EC0UW87 (L19201) by 93 bp. NOTE: An update was  
submitted on 25-OCT-1993, reflecting a correction to the hemE

sequence.

FEATURES

source

Location/Qualifiers

1. 176195

/organism="Escherichia coli"

/mol\_type="unassigned DNA"

/strain="K-12"

/sub\_strain="MG1655"

/db\_xref="taxon:562"

/note="This sequence comprises the following lambda

clones: EC18-126, EC14-54, EC21-52, EC17-24, EC17-203,

EC19-105, EC18-208, EC18-110, EC17-200, EC18-33, EC30-262,

EC30-41, EC30K637-2a, EC30K637-5, EC27-975, EC22-27;

M13mp19 or Janus vectors were used for subcloning"

misc\_feature  
gene  
CDS  
misc\_feature  
misc\_feature  
misc\_feature  
misc\_difference  
terminator  
CDS  
misc\_difference  
misc\_difference  
CDS  
promoter  
misc\_structure  
repeat\_region  
misc\_structure

<1..1737  
/notes="corresponds to M21516; ECOKATGA (1225..2805)"  
1..1257  
/gene="katG"  
CDS  
1..1257  
/gene="katG"  
/EC\_number="1.11.1.6"  
/standard\_name="catalase HPI; catalase-peroxide; catalase"  
/notes="CG Site No. 14983"  
/codon\_start=1  
/transl\_table=11  
/product="catalase hydroperoxidase I"  
/protein\_id="AAC43048.1"  
/db\_xref="GI:396289"  
/translation="AITSGLEVVWTTQTQWSNYFENLFKYEWVQTRSPAGAIQPEA  
VPAEIIIPDPFPPSKRKPTMLVTDLTRDFEFKISRRLNDPOAFNEAFARXWFK  
LTHRMGPKSRVIGPVPKEDLIWDPPLPQPIYNPTEODIIDLKEAISGSLVSELY  
SVNASASTFRGDKRGKANGARKALMPORDWDVNAARALPVLKEIKQESGKSELV  
DIIVAGVVGVKAAAGLSHVPPAPGRVDARQDQDIEMFELLEPIADGFRNYRA  
RLDVSTESLIDKAQQLTAPMTALVGMRLGANFDGSKNGVFTDRVGVLNDF  
FVNLDMRYEMKATDSKELFEGRDRETGEVKTASRADLVFGSNSVLRAVAEYVASS  
DAHEKPVKDFVAWVKVMDLRFDLL"  
CDS  
1..161  
/gene="katG"  
/notes="corresponds to lambda clone EC18-126"  
1..93  
/gene="katG"  
/note="93 bp overlap with end of L19201 (ECOWU87)"  
156..15760  
/note="corresponds to lambda clone EC14-54"  
misc\_difference  
938  
/gene="katG"  
/note="G in M21516; C here"  
1266..1305  
/note="putative"  
CDS  
1317..2255  
/note="similar to Desulfurolobus ambivalens hypoth. 28.3  
kDa protein in sor 3' region"  
/codon\_start=1  
/transl\_table=11  
/label=ORF\_0312  
/protein\_id="AAC43049.1"  
/db\_xref="GI:409786"  
/translation="MGERKLHTGSLMSAAGKNPLAISGLVLTILWSYMIFFKQVT  
SYIGAFDFALRCIFGLVLFVLLRGRMPTPEKTYLTAIALIQTQCMVGLAQMAL  
VSGAGKVAIIYSYTWPFMWVIFAAFLGRLRGQYFALLIAFGLFLVLQWDLDFS  
SNKSNAILSGVSGASAIKRLYARHPVDLLSLTSQWMLYALVMSVVALVPQ  
REIDQPTVFWALYSAILATALWSMLFVLKLPASPASIASLSTILAVPVCGLFSWML  
LGENPFGVSGSGLVILVALALVSRKKKEAVSVKRI"  
misc\_difference  
1429..1432  
/note="TTT in M21516; TTTT here"  
CDS  
1720  
/note="C in M21516; G here"  
complement(2282..2899)  
/codon\_start=1  
/transl\_table=11  
/label=ORF\_f205  
/protein\_id="AAC43050.1"  
/db\_xref="GI:396291"  
/translation="WKASLALLSLTATFTSHLSKSPAVPPTVQIQANTNLAIADGAR  
QQIGSTLYDPAYQLTYPGGDVQPGVCSDVIRALRSQKVDLQKLVHEDMAKNFA  
EYQWKLRPDSDNIDHRRVPLETWFSRHDKTRPTSKNPSDYQAGDIVSWRLDGLA  
HIGVYDGFARDGTPLVTHNICAGAEEDVLFNRMVGHRYFVK"  
promoter  
complement(2886..2915)  
/note="promoter matrix score of 42; putative; within ORF  
f205, which would suggest alternate start codon"  
2912..3053  
/note="predicted bend of 81 degrees"  
misc\_structure  
3031..3064  
/standard\_name="REP; repetitive extragenic palindromic  
element"  
/note="contains 1 REP sequence"  
misc\_structure  
3090..3219

gene  
CDS  
misc\_structure  
CDS  
promoter  
CDS  
CDS  
promoter  
misc\_feature  
terminator

/note="predicted bend of 75 degrees"  
complement(3174..4316)  
/gene="gldA"  
complement(3174..4316)  
/gene="gldA"  
/note="similar to Bacillus stearothermophilus glycerol  
dehydrogenase"  
/codon\_start=1  
/transl\_table=11  
/label=ORF\_f380  
/product="Glycerol dehydrogenase"  
/protein\_id="AAC43051.1"  
/db\_xref="GI:396292"  
/translation="NPHLALLSKGAIMDRIIIOSPGKIQAGADVNRILRGYLKPLXER  
WLVGDKFVLGPAQSTVKEFSKDGALVVEIAPFGGCSONEIDRLRGIARTACGAIL  
GIGGKTLDTAKALAHFMGVVALIARTIASTDAPCSALSVIYTDGEFDRYLLPNP  
NNVIYDTKIVAGAPARLLAAGIGDALTWFEARACSRGATWAGCGTQAALALAE  
CYNTHLEEGEKAMLAEEQHVTPALERVIEANTYLSGVFGSGGLAAAHAVHGLTAI  
PDAPHYHGEKVAFGTLTOLVENAPVEIETVAALSHAVGLPITITLAQLDIKEDVPAK  
MRIVAAACAEGEETHNMPGGATPDQVYAALLVADQYQGRFLQEW"  
4022..4156  
/note="predicted bend of 75 degrees"  
complement(4288..4950)  
/note="similar to Bacillus subtilis hypoth. 20 kDa  
protein, in tsr 3' region"  
/codon\_start=1  
/transl\_table=11  
/label=ORF\_f220  
/protein\_id="AAC43052.1"  
/db\_xref="GI:396293"  
/translation="MELYLDTANVAEVLRLARIFPIAGVTTNPSTIAASKESIWEVLP  
RLQKAIGDEGILFAOTMSRDAQMVEEAKRLDAIPGVVKIPVTSEGLAAIKLKE  
GITLTGAVYSAOGLLALAGAKYVAVVNEVDAGGDIETVOKLTLLLENHAPES  
MYLAASFYTPROALDCLLAGCESITPLDVAQOQMLNTPAVESAIEFEDHWNAAFST  
HL"  
promoter  
complement(4922..4949)  
/note="promoter matrix score of 47; putative; within ORF  
f220, which would suggest alternate start codon"  
CDS  
complement(4962..7097)  
/note="similar to phosphotransferase system enzyme I"  
/codon\_start=1  
/transl\_table=11  
/label=ORF\_f711  
/protein\_id="AAC43053.1"  
/db\_xref="GI:409787"  
/translation="MCSSSAGGILTPISDLNLALNLPAAKGVDAEQSALENGLTV  
LKNIFRLDSDGATSAILEAHRSLAGDTSLEHLLAGVSAGLSACAAIVASANHFCE  
EFSRSSYQLBERALDVRDVCQLLQIYGEORFPAPGKLTQPAICMADELTPSQALV  
LDKNHLGILLKSGGTTSHTVILARSFMTLVGVDDIDALTQWQQQTIYIDGNAGAI  
VEPGEAVARYQOEARVQDALRQORVMTLQOARTADGIRIBIAANIISVEAQAAPF  
NGAEGVGLFRTEMLYMDRTSAPGESELYNIFCOALESANGRSIIIVTMDISGDKPVDY  
LNIPEANPFLGYRAVRIYEYASLFTQLRSILRASAHGSLKIMPMISSMBEILLVY  
KEKLAERQQRLNEHIPDEKIQGLGMLEVSPVMTIDQCCBEIDFFSIGSNDLTQYL  
LAVORDNAKVTNRHNSLPAFLDALDYAVQVHRQGWIGLCELGAGKSVLPDLVGL  
GLDELSMSPISPAAKARMAQDLSRECKLLNQAMACRTSLSEVHLLAQFRMTQDAP  
LVTAEICITLESDPKRSKEVLTGMNTDNLLAGRCYRPRLEADLWAEAVFSTGLFSF  
AIPHSKSEHIEQSTISVARLOAPVPGDDEAQFIIMLTLNKHAAGQDHMRIFSRLLAR  
IMHSEFRNALVNAASADALIASLLQHELEL"  
7099..7428  
/codon\_start=1  
/transl\_table=11  
/label=ORF\_0109  
/protein\_id="AAC43054.1"  
/db\_xref="GI:409788"  
/translation="MRARIICGFRLVSETSGSSSSDLTASGASQMGNSRSHWLKRI  
WCACSSAPEISNWLSPASVSPISARALLALPCGCAGYATQPFVMKYYTAFQRLG  
PDAHVR"  
complement(7108..7136)  
/note="promoter matrix score of 46; putative"  
7200..25839  
/note="corresponds to lambda clone EC21-52"  
7578..7699  
/note="putative"

CDS	7771..8850	/note="similar to phosphotransferase system enzyme II"
	/codon_start=1	
	/transl_table=11	
	/label=ORF_0359	
	/protein_id="AAC43055.1"	
	/db_xref="GI:396296"	
	/translation="MNELVQITLKNTRHQLMTGVSHMIPFVVSGGILLAVSVMLYKGKA	
	VPDAVDANLKKLPDGVAGLTLMVPLAAVIGYSIAERSALAPCAICAWGVNSFGAG	
	FFGALIGKIGIVVHYLKKLPVHKLVSMPRIPIPIVGLTITAGIMWGLGEPGAG	
	LTNSLTQWLQMGQGSIVIMGLMPLAFMGGVNVKVAYAFMLDICVAQGYIVTVVAI	
	AAVGICIPPLGMGLATLIGRKNFSABERETGKAALVMGCVGTGTGAIPFAAADPLRVI	
Query Match	72.8%; Score 498; DB 15; Length 176195;	
Best Local Similarity	100.0%; Pred. No. 3.4e-143;	
Matches 498; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
Qy	187 ATGTCACACCCCGGTTAAACCAACTGCGTCGCTATTGTAAGAGATCCCTGCC	246
Db	117748 ATGTCACACCCCGGTTAAACCAACTGCGTCGCTATTGTAAGAGATCCCTGCC	117807
Qy	247 CTGATCCGCAACTGCTCGACTGGCTGTTCTGGAGGATTCATGACAAAAACGTTTGA	306
Db	117808 CTGATCCGCAACTGCTCGACTGGCTGTTCTGGAGGATTCATGACAAAAACGTTTGA	117867
Qy	307 CAGCAGGGAAAAACGGTAAGCGTGACGATGATCCGCGAAGGTTTGTCCGACGAGATCAA	366
Db	117868 CAGCAGGGAAAAACGGTAAGCGTGACGATGATCCGCGAAGGTTTGTCCGACGAGATCAA	117927
Qy	367 ATCCCCGAAGAACTGCCCTGCTGCCGAAGAGTCTCGTTACTGGTTACGTGAATTTTG	426
Db	117928 ATCCCCGAAGAACTGCCCTGCTGCCGAAGAGTCTCGTTACTGGTTACGTGAATTTTG	117987
Qy	427 TTATGTGCCGATGGTGAAACCGTGCTGCCGGTCGTACCGTCGTCTGTGTCAACGTTA	486
Db	117988 TTATGTGCCGATGGTGAAACCGTGCTGCCGGTCGTACCGTCGTCTGTGTCAACGTTA	118047
Qy	487 AGCGGCGGAGCTGGCGTTACAAAAATGGGTAAACCGCGTTAGGACGCTATCTGTC	546
Db	118048 AGCGGCGGAGCTGGCGTTACAAAAATGGGTAAACCGCGCTTAGGACGCTATCTGTC	118107
Qy	547 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGCCGCTGATGCCGGCTGTGGGGG	606
Db	118108 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGCCGCTGATGCCGGCTGTGGGGG	118167
Qy	607 CGACGTTCCCGCTCGCATTAAGCGGTAAACCGCTGTTGTCTAAACAGAACTGTTTTACCG	666
Db	118168 CGACGTTCCCGCTCGCATTAAGCGGTAAACCGCTGTTGTCTAAACAGAACTGTTTTACCG	118227
Qy	667 CGGTCACCGTTGTACTAA	684
Db	118228 CGGTCACCGTTGTACTAA	118245

Search completed: May 30, 2006, 02:12:44  
Job time : 4102 secs

GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.  
OM nucleic - nucleic search, using sw model  
Run on: May 29, 2006, 10:44:22 ; Search time 322 Seconds  
(without alignments)  
3974.653 Million cell updates/sec

Title: US-10-718-311-7  
Perfect score: 684  
Sequence: 1 atggcttcctctgtcatttc.....cggcgtaacctgtactaa 684  
Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 1403666 seqs, 935554401 residues  
Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA:\*  
1: /EMC\_Celerra\_SIDS3/ptodata/2/ina/1 COMB.seq:\*  
2: /EMC\_Celerra\_SIDS3/ptodata/2/ina/5 COMB.seq:\*  
3: /EMC\_Celerra\_SIDS3/ptodata/2/ina/6A COMB.seq:\*  
4: /EMC\_Celerra\_SIDS3/ptodata/2/ina/6B COMB.seq:\*  
5: /EMC\_Celerra\_SIDS3/ptodata/2/ina/7 COMB.seq:\*  
6: /EMC\_Celerra\_SIDS3/ptodata/2/ina/H COMB.seq:\*  
7: /EMC\_Celerra\_SIDS3/ptodata/2/ina/PCTUS COMB.seq:\*  
8: /EMC\_Celerra\_SIDS3/ptodata/2/ina/PP COMB.seq:\*  
9: /EMC\_Celerra\_SIDS3/ptodata/2/ina/RE COMB.seq:\*  
10: /EMC\_Celerra\_SIDS3/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description
1	684	100.0	US-09-896-866B-16 Sequence 16, Appl
2	684	100.0	US-09-855-341-7 Sequence 7, Appl
3	498	72.8	US-09-064-693A-24 Sequence 24, Appl
4	498	72.8	US-09-064-693A-25 Sequence 25, Appl
5	495	72.4	US-09-896-866B-12 Sequence 12, Appl
6	495	72.4	US-09-855-341-3 Sequence 3, Appl
7	282	41.2	US-09-489-039A-1234 Sequence 1234, Ap
8	140.8	20.6	US-09-839-477-3 Sequence 3, Appl
9	140.8	20.6	US-09-495-797-41 Sequence 41, Appl
10	133.2	19.5	US-08-152-483B-8 Sequence 8, Appl
11	130.6	19.1	US-08-095-726-12 Sequence 12, Appl
12	130.6	19.1	US-08-096-043-12 Sequence 12, Appl
13	130.6	19.1	US-08-093-577-8 Sequence 8, Appl
14	130.6	19.1	US-08-331-004A-6 Sequence 6, Appl
15	130.6	19.1	US-08-096-623A-12 Sequence 12, Appl
16	130.6	19.1	PCT-US95-13937A-6 Sequence 6, Appl
17	126.6	18.5	US-09-543-681A-3122 Sequence 3122, Ap
18	97.2	14.2	US-09-839-477-5 Sequence 5, Appl
19	77.2	11.3	US-08-469-202-15 Sequence 15, Appl
20	77.2	11.3	US-08-469-202-17 Sequence 17, Appl
21	77.2	11.3	US-08-469-202-19 Sequence 19, Appl
22	77.2	11.3	US-08-484-434C-15 Sequence 15, Appl
23	77.2	11.3	US-08-484-434C-18 Sequence 18, Appl

ALIGNMENTS

RESULT 1  
US-09-896-866B-16  
; Sequence 16, Application US/09896866B  
; Patent No. 6642036  
; GENERAL INFORMATION:  
; APPLICANT: Flint, Dennis  
; APPLICANT: Meyer, Knut  
; APPLICANT: Viitanen, Paul  
; TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugate  
; FILE REFERENCE: BC1034 US NA  
; CURRENT APPLICATION NUMBER: US/09/896,866B  
; CURRENT FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: 60/216,615  
; PRIOR FILING DATE: 2000-07-07  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 16  
; LENGTH: 684  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; NAME/KEY: misc feature  
; OTHER INFORMATION: open reading frame of the chloroplast-targeted CPL fusion protei  
US-09-896-866B-16

Query Match	100.0%;	Score 684;	DB 3;	Length 684;
Best Local Similarity	100.0%;	Pred. No. 1.1e-230;		
Matches 684;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGSGCTTCCTCTGTCATTTCTTCAGCAGCTGTTGCCACACGACGAGCAATGTTACACAGCT	60	
Db	1	ATGSGCTTCCTCTGTCATTTCTTCAGCAGCTGTTGCCACACGACGAGCAATGTTACACAGCT	60	
QY	61	AGCATGGTTGACACCTTTCACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTCACAAAGAAG	120	
Db	61	AGCATGGTTGACACCTTTCACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTCACAAAGAAG	120	
QY	121	CAAAACCTTGACATCACTTCCATTTGTCAGCAATGTTGGAAGAGTTAGTGCATGCAGGTG	180	
Db	121	CAAAACCTTGACATCACTTCCATTTGTCAGCAATGTTGGAAGAGTTAGTGCATGCAGGTG	180	
QY	181	TGGCATATGTCACACCCCGGTTAAGCAACTCGCGTGCCTATTGTTAAAGAGATC	240	
Db	181	TGGCATATGTCACACCCCGGTTAAGCAACTCGCGTGCCTATTGTTAAAGAGATC	240	

241	Qy	CTGCCCTGGATCCGGA	CTGCTCGACTGGCTGTGTCGAGGATTCATGACAAACGT	300
241	Db	CTGCCCTGGATCCGGA	CTGCTCGACTGGCTGTGTCGAGGATTCATGACAAACGT	300
301	Qy	TTTGAACACGACGGAAAA	CGGTAGCTGACGATGATCCGCGAAGGGTTTGTGCGAGCAG	360
301	Db	TTTGAACACGACGGAAAA	CGGTAGCTGACGATGATCCGCGAAGGGTTTGTGCGAGCAG	360
361	Qy	AATGAATCCCGAAGAA	CTGCCGTGTCGCGAAGAGTCTCGTTACTGTTTACGTGAA	420
361	Db	AATGAATCCCGAAGAA	CTGCCGTGTCGCGAAGAGTCTCGTTACTGTTTACGTGAA	420
421	Qy	ATTTTGTATATGCGGAT	GTCGACCGTGGCTTCGCGTCGTACCGTCTGTCGTGTC	480
421	Db	ATTTTGTATATGCGGAT	GTCGACCGTGGCTTCGCGTCGTACCGTCTGTCGTGTC	480
481	Qy	ACGTTAAGCGGCGCGAG	CTGCGCTTACAAAAATTGGGTAAAAACGCCGTTAGGACGCTAT	540
481	Db	ACGTTAAGCGGCGCGAG	CTGCGCTTACAAAAATTGGGTAAAAACGCCGTTAGGACGCTAT	540
541	Qy	CTGTTACATCATCGACAT	TAAACCGGACCTTTATTCAGATAGCCCGTGTATGCCGGGCTG	600
541	Db	CTGTTACATCATCGACAT	TAAACCGGACCTTTATTCAGATAGCCCGTGTATGCCGGGCTG	600
601	Qy	TGGGGCGACGCTCCCG	CGCTTCGATTAGCGGTAAACCGCTGTGCTGTAAACAGAACTGT	660
601	Db	TGGGGCGACGCTCCCG	CGCTTCGATTAGCGGTAAACCGCTGTGCTGTAAACAGAACTGT	660
661	Qy	TTACCGGCGTCAACG	TTGTACTAA	684
661	Db	TTACCGGCGTCAACG	TTGTACTAA	684

## RESULT 2

```

US-09-855-341-7
; Sequence 7, Application US/09855341
; Patent No. 6683231
; GENERAL INFORMATION:
; APPLICANT: VIITANEN, PAUL V.
; APPLICANT: MEYER, KNUIT
; APPLICANT: VAN DYK, DREW
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID
; TITLE OF INVENTION: IN GREEN PLANTS
; FILE REFERENCE: BC1015 US NA
; CURRENT APPLICATION NUMBER: US/09/855,341
; CURRENT FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 7
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:synthetic CPL
US-09-855-341-7

```

181	Db		TGGCATATGTCACACCCCGCGTTTAA	CGCAACTGCGTTCGCGTTCGCTATTGTTAAAGAGATC	240
241	Qy		CCTCCCTGGATCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTC	CAATGACAAAACGT	300
241	Db		CCTGCCCTGGATCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTC	CAATGACAAAACGT	300
301	Qy		TTTGAAACAGCAGGGAATAAACCGTAAAGCGTGCAGATGCCGGAAGGGTTT	GTGCGAGCAG	360
301	Db		TTTGAAACAGCAGGGAATAAACCGTAAAGCGTGCAGATGCCGGAAGGGTTT	GTGCGAGCAG	360
361	Qy		AATCAATCCCCGAGAACTGCCGCTGCTCCGGAAGAGTCTCGTACTGGTTAC	GTGTACGTGAA	420
361	Db		AATCAATCCCCGAGAACTGCCGCTGCTCCGGAAGAGTCTCGTACTGGTTAC	GTGTACGTGAA	420
421	Qy		ATTTTGTTATGTGCCGATGGTGAAACCGTGGCTTGCCGCTCGTACCGT	CGCTTCTGTGTCA	480
421	Db		ATTTTGTTATGTGCCGATGGTGAAACCGTGGCTTGCCGCTCGTACCGT	CGCTTCTGTGTCA	480
481	Qy		ACGTTAAGCGGCGCGAGCTCGCGTTACAAAAATTGGGTAAACCGCGT	TAGGACGCTAT	540
481	Db		ACGTTAAGCGGCGCGAGCTCGCGTTACAAAAATTGGGTAAACCGCGT	TAGGACGCTAT	540
541	Qy		CTGTTTCATCATCGACATTTAACCGGAGCTTTATTGAGATAGGCGGTGAT	CGCCGGCTG	600
541	Db		CTGTTTCATCATCGACATTTAACCGGAGCTTTATTGAGATAGGCGGTGAT	CGCCGGCTG	600
601	Qy		TGGGGGCGACGTTCCCGCTCCGATTAAAGCGGTAAACCGCTGTTGCTAA	CAGAACTGTTT	660
601	Db		TGGGGGCGACGTTCCCGCTCCGATTAAAGCGGTAAACCGCTGTTGCTAA	CAGAACTGTTT	660
661	Qy		TTACCGGGCTCACCGTTGTACTAA	684	
661	Db		TTACCGGGCTCACCGTTGTACTAA	684	

### RESULT 3

US-09-064-693A-24  
; Sequence 24, Application US/09064693A  
; Patent No. 6210937  
; GENERAL INFORMATION:  
; APPLICANT: Ward, Thomas E.  
; TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY  
; TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION  
; TITLE OF INVENTION: OF A SPECIFIC PLASTICS PRECURSOR  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: W. Gary Goodson  
; ADDRESSEE: INEL--Lockheed Martin Idaho  
; ADDRESSEE: Technologies Co.  
; STREET: P.O. Box 1625  
; CITY: Idaho Falls  
; STATE: Idaho  
; COUNTRY: USA  
; ZIP: 83415-3810  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: Toshiba Satellite Pro T2150CDS  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: Word Perfect 7.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/064,693A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: W. Gary Goodson  
; REGISTRATION NUMBER: 22,387  
; REFERENCE/DOCKET NUMBER: LIT-PI-296  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (208)526-9469

TELEFAX: (208)526-8339  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 854 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-064-693A-24

Query Match 72.8%; Score 498; DB 3; Length 854;  
Best Local Similarity 100.0%; Pred. No. 6.2e-165; Mismatches 0; Indels 0; Gaps 0;  
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 187 ATGTACACCCCGGTTAAACGCAACTCGCTGCGCTGCGCTATTTGTAAGAGATCCCTGCC 246  
DB 321 ATGTACACACCCCGGTTAAACGCAACTCGCTGCGCTATTTGTAAGAGATCCCTGCC 380  
QY 247 CTGGATCCGCAACTGCTGACCTGGCTGTTGCTGGAGATTCATGACAAAACGTTTGA 306  
DB 381 CTGGATCCGCAACTGCTGACCTGGCTGTTGCTGGAGATTCATGACAAAACGTTTGA 440  
QY 307 CAGCAGGGGAAAAACGTTAGCGTACGATCATCCGGAAGGGTTTCTCGAGCAGATGAA 366  
DB 441 CAGCAGGGGAAAAACGTTAGCGTACGATCATCCGGAAGGGTTTCTCGAGCAGATGAA 500  
QY 367 ATCCCGGAAGAACTGCGCTGCTGCGGAAAGAGTCTCTGTTACTGTTAGCTGAAATTTG 426  
DB 501 ATCCCGGAAGAACTGCGCTGCTGCGGAAAGAGTCTCTGTTACTGTTAGCTGAAATTTG 560  
QY 427 TTATGTCCGATGTTGAAACGTTGGCTGTCGGGTGACCGTCTTCTGTGTCAACGTTA 486  
DB 561 TTATGTCCGATGTTGAAACGTTGGCTGTCGGGTGACCGTCTTCTGTGTCAACGTTA 620  
QY 487 AGCGGGCGGAGTGGCGTTACAAAATTTGGTTAAACCGCTTAGGAGCGTATCTGTTTC 546  
DB 621 AGCGGGCGGAGTGGCGTTACAAAATTTGGTTAAACCGCTTAGGAGCGTATCTGTTTC 680  
QY 547 ACATCATCGACATTAACCGGAGCTTTATTGAGATAGCGCGTGATGCCGGCTGTGGGG 606  
DB 681 ACATCATCGACATTAACCGGAGCTTTATTGAGATAGCGCGTGATGCCGGCTGTGGGG 740  
QY 607 CGACGTTCCCGCTGGGATTAAGCGGTAACCGCTGTTGCTTAACAGAACTGTTTTACCG 666  
DB 741 CGACGTTCCCGCTGGGATTAAGCGGTAACCGCTGTTGCTTAACAGAACTGTTTTACCG 800  
QY 667 GCGTCACCGTTGACTAA 684  
DB 801 GCGTCACCGTTGACTAA 818

RESULT 4  
US-09-064-693A-25  
Sequence 25, Application US/09064693A  
Patent No. 6210937  
GENERAL INFORMATION:  
APPLICANT: Ward, Thomas E.  
TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY  
ENGINEERED BACTERIA FOR PRODUCTION  
OF A SPECIFIC PLASTICS PRECURSOR  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSER: W. Gary Goodson  
ADDRESSEE: INEEL--Lockheed Martin Idaho  
ADDRESSEE: Technologies Co.  
STREET: P.O. Box 1625  
CITY: Idaho Falls  
STATE: Idaho  
COUNTRY: USA  
ZIP: 83415-3810  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: Toshiba Satellite Pro T2150CDS  
OPERATING SYSTEM: Windows95

SOFTWARE: Word Perfect 7.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/064,693A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: W. Gary Goodson  
REGISTRATION NUMBER: 22,387  
REFERENCE/DOCKET NUMBER: LIT-PI-296  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (208)526-8339  
TELEFAX: (208)526-8339  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 664 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-064-693A-25  
Query Match 72.8%; Score 498; DB 3; Length 664;  
Best Local Similarity 100.0%; Pred. No. 2.2e-164; Mismatches 0; Indels 0; Gaps 0;  
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 187 ATGTACACCCCGGTTAAACGCAACTCGCTGCGCTGCGCTATTTGTAAGAGATCCCTGCC 246  
DB 321 ATGTACACCCCGGTTAAACGCAACTCGCTGCGCTGCGCTATTTGTAAGAGATCCCTGCC 380  
QY 247 CTGGATCCGCAACTGCTGACCTGGCTGTTGCTGGAGATTCATGACAAAACGTTTGA 306  
DB 381 CTGGATCCGCAACTGCTGACCTGGCTGTTGCTGGAGATTCATGACAAAACGTTTGA 440  
QY 307 CAGCAGGGGAAAAACGTTAGCGTACGATCATCCGGAAGGGTTTCTCGAGCAGATGAA 366  
DB 441 CAGCAGGGGAAAAACGTTAGCGTACGATCATCCGGAAGGGTTTCTCGAGCAGATGAA 500  
QY 367 ATCCCGGAAGAACTGCGCTGCTGCGGAAAGAGTCTCTGTTACTGTTAGCTGAAATTTG 426  
DB 501 ATCCCGGAAGAACTGCGCTGCTGCGGAAAGAGTCTCTGTTACTGTTAGCTGAAATTTG 560  
QY 427 TTATGTCCGATGTTGAAACGTTGGCTGTCGGGTGACCGTCTTCTGTGTCAACGTTA 486  
DB 561 TTATGTCCGATGTTGAAACGTTGGCTGTCGGGTGACCGTCTTCTGTGTCAACGTTA 620  
QY 487 AGCGGGCGGAGTGGCGTTACAAAATTTGGTTAAACCGCTTAGGAGCGTATCTGTTTC 546  
DB 621 AGCGGGCGGAGTGGCGTTACAAAATTTGGTTAAACCGCTTAGGAGCGTATCTGTTTC 680  
QY 547 ACATCATCGACATTAACCGGAGCTTTATTGAGATAGCGCGTGATGCCGGCTGTGGGG 606  
DB 681 ACATCATCGACATTAACCGGAGCTTTATTGAGATAGCGCGTGATGCCGGCTGTGGGG 740  
QY 607 CGACGTTCCCGCTGGGATTAAGCGGTAACCGCTGTTGCTTAACAGAACTGTTTTACCG 666  
DB 741 CGACGTTCCCGCTGGGATTAAGCGGTAACCGCTGTTGCTTAACAGAACTGTTTTACCG 800  
QY 667 GCGTCACCGTTGACTAA 684  
DB 801 GCGTCACCGTTGACTAA 818

RESULT 5  
US-09-896-866B-12  
Sequence 12, Application US/09896866B  
Patent No. 6642036  
GENERAL INFORMATION:  
APPLICANT: Flint, Dennis  
APPLICANT: Meyer, Knut  
APPLICANT: Viitanen, Paul  
TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugate:



```
; TITLE OF INVENTION: Benzolic Acid Glucosides
; FILE REFERENCE: BC1034 US NA
; CURRENT APPLICATION NUMBER: US/09/896,866B
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/216,615
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 12
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-896-866B-12

Query Match      72.4%; Score 495; DB 3; Length 495;
Best Local Similarity 100.0%; Pred. No. 5.1e-164;
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGGCTGCGCTATTGTAAAGAGATCCCTGCC 246
DB 1 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGGCTGCGCTATTGTAAAGAGATCCCTGCC 60

QY 247 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAAACGTTTTGAA 306
DB 61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAAACGTTTTGAA 120

QY 307 CAGCAGGGAAAAACCGTAAGCGTGACGATGATCCGCGAAGGGTTGTGCGAGCAGATGAA 366
DB 121 CAGCAGGGAAAAACCGTAAGCGTGACGATGATCCGCGAAGGGTTGTGCGAGCAGATGAA 180

QY 367 ATCCCGGAAGAACTGCCGCTGCTGCCGAAAGAGTCTGTTACTGTTACGTGGAATTTTG 426
DB 181 ATCCCGGAAGAACTGCCGCTGCTGCCGAAAGAGTCTGTTACTGTTACGTGGAATTTTG 240

QY 427 TTATGTGCGCGATGTTGAACCGTGGCTTGCCCGTGGTACCGTCTGTTCTGTGTCAACGTTA 486
DB 241 TTATGTGCGCGATGTTGAACCGTGGCTTGCCCGTGGTACCGTCTGTTCTGTGTCAACGTTA 300

QY 487 AGCGGGCCGAGCTGGCGGTTACAAAAATTTGGTAAAAACGCCGTTAGGACGCTATCTGTTTC 546
DB 301 AGCGGGCCGAGCTGGCGGTTACAAAAATTTGGTAAAAACGCCGTTAGGACGCTATCTGTTTC 360

QY 547 ACATCATCGACATTAACCCCGGACTTTATTGAGATAGGCGGTGATGCCGGCTGTGGGG 606
DB 361 ACATCATCGACATTAACCCCGGACTTTATTGAGATAGGCGGTGATGCCGGCTGTGGGG 420

QY 607 CGAGTTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTACCG 666
DB 421 CGAGTTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTACCG 480

QY 667 GCGTCACCGTTGTAC 681
DB 481 GCGTCACCGTTGTAC 495
```

```
RESULT 6
US-09-855-341-3
; Sequence 3, Application US/09855341
; Patent No. 6683231
; GENERAL INFORMATION:
; APPLICANT: VIITANEN, PAUL V.
; APPLICANT: MEYER, KNUT
; APPLICANT: VAN DYK, DREW
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID
; FILE REFERENCE: BC1015 US NA
; CURRENT APPLICATION NUMBER: US/09/855,341
; CURRENT FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 3
; LENGTH: 495
; TYPE: DNA
```

```
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism:E. coli
US-09-855-341-3

Query Match      72.4%; Score 495; DB 3; Length 495;
Best Local Similarity 100.0%; Pred. No. 5.1e-164;
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGGCTGCGCTATTGTAAAGAGATCCCTGCC 246
DB 1 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGGCTGCGCTATTGTAAAGAGATCCCTGCC 60

QY 247 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAAACGTTTTGAA 306
DB 61 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAAACGTTTTGAA 120

QY 307 CAGCAGGGAAAAACCGTAAGCGTGACGATGATCCGCGAAGGGTTGTGCGAGCAGATGAA 366
DB 121 CAGCAGGGAAAAACCGTAAGCGTGACGATGATCCGCGAAGGGTTGTGCGAGCAGATGAA 180

QY 367 ATCCCGGAAGAACTGCCGCTGCTGCCGAAAGAGTCTGTTACTGTTACGTGGAATTTTG 426
DB 181 ATCCCGGAAGAACTGCCGCTGCTGCCGAAAGAGTCTGTTACTGTTACGTGGAATTTTG 240

QY 427 TTATGTGCGCGATGTTGAACCGTGGCTTGCCCGTGGTACCGTCTGTTCTGTGTCAACGTTA 486
DB 241 TTATGTGCGCGATGTTGAACCGTGGCTTGCCCGTGGTACCGTCTGTTCTGTGTCAACGTTA 300

QY 487 AGCGGGCCGAGCTGGCGGTTACAAAAATTTGGTAAAAACGCCGTTAGGACGCTATCTGTTTC 546
DB 301 AGCGGGCCGAGCTGGCGGTTACAAAAATTTGGTAAAAACGCCGTTAGGACGCTATCTGTTTC 360

QY 547 ACATCATCGACATTAACCCCGGACTTTATTGAGATAGGCGGTGATGCCGGCTGTGGGG 606
DB 361 ACATCATCGACATTAACCCCGGACTTTATTGAGATAGGCGGTGATGCCGGCTGTGGGG 420

QY 607 CGAGTTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTACCG 666
DB 421 CGAGTTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAAACAGAACTGTTTTACCG 480

QY 667 GCGTCACCGTTGTAC 681
DB 481 GCGTCACCGTTGTAC 495
```

```
RESULT 7
US-09-489-039A-1234
; Sequence 1234, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1234
; LENGTH: 645
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1234
```

```
Query Match      41.2%; Score 282; DB 3; Length 645;
Best Local Similarity 72.9%; Pred. No. 1e-88;
Matches 363; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 187 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGCGCTATTGTAAAGAGATCCCTGCC 246
DB 148 ATGTCACATCTCGCGCTTACGCGACTGCGTGGCTGCGCTATTTTGGCCGTTATGCCATCC 207
```



```

RESULT 9
US-09-435-797-41
; Sequence 41, Application US/09495797
; Patent No. 6369296
; GENERAL INFORMATION:
; APPLICANT: Ratcliff, Frank G
; APPLICANT: Martin-Hernandez, Ana M
; APPLICANT: Bulcombe, David C
; TITLE OF INVENTION: Viral Vectors
; FILE REFERENCE: Newduth 43, 047
; CURRENT APPLICATION NUMBER: US/09/495,797
; CURRENT FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 41
; LENGTH: 499
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: N. benthamiana rubisco
US-09-435-797-41

```

Query Match	20.6%	Score 140.8;	DB 3;	Length 499;
Best Local Similarity	85.3%	Pred. No. 6.3e-39;		
Matches 157; Conservative	0;	Mismatches 27;	Indels 0;	Gaps 0

Qy	1	ATGGCTTCCCTGTGTCATTTCTTTCAGCAGCTGTTGCCACACGACGAAATGTTACACAAGCT	60
Db	8	ATGGCTTCCCTCAGTTCTTTCTCTGTCAGCAGTTGCCACCCGACGAATGTTGCTCAAAGCT	67
Qy	61	AGCATGGTTGCACCTTTTCACCTGGTCTCAAAATCTTTCAGCCACTTTCCCTGTTTCAAAAGAAG	120
Db	68	AACATGGTTGCACCTTTTCACCTGGCTTTAAGTCAGCTGCTCATTCCCTGTTTCAAGGAAG	127
Qy	121	CAAAACCTTGACATCACCATTCCATTGCTAGCAATGGTGGAAAGAGTTAGCTGCATGCAAGGTG	180
Db	128	CAAAACCTTGACATCACCATTCCATTGCTGCAACGAACGGCGGAAGAGTCAATGTCATGCAGGTG	187
Qy	181	TGGC	184
Db	188	TGGC	191

```

1  TITLE OF INVENTION: PROTEIN FUSION ENZYME
2
3  NUMBER OF SEQUENCES: 9
4
5  CORRESPONDENCE ADDRESS:
6
7  ADDRESSEE: Pennie & Edmonds
8  STREET: 2730 Sand Hill Road
9  CITY: Menlo Park
10 STATE: California
11
12 COUNTRY: U.S.A.
13
14 ZIP: 94025
15
16 COMPUTER READABLE FORM:
17
18 MEDIUM TYPE: Floppy disk
19
20 COMPUTER: IBM PC compatible
21
22 OPERATING SYSTEM: PC-DOS/MS-DOS
23
24 SOFTWARE: Patent in Release #1.0,
25 SOFTWARE: Version #1.25

```

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/152,483B
; FILING DATE: No. 5529909ember 12, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 7/857,602
; FILING DATE: March 30, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 923,692
; FILING DATE: July 31, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 600,244
; FILING DATE: October 22, 1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 641,617
; FILING DATE: January 16, 1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 737,899
; FILING DATE: July 26, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P.
; REGISTRATION NUMBER: 25,227
; REFERENCE/DOCKET NUMBER: BIOG-20240/8129-040
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 854-3660
; TELEFAX: (415) 854-3694
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1442
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM:
; IMMEDIATE SOURCE:
; CLONE:
; FEATURE:
;
US-08-152-483B-8

Query Match 19.5%; Score 133.2; DB 2; Length 1442;
Best Local Similarity 84.3%; Pred. No. 5.9e-36;
Matches 150; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 1 ATGGCTTCCTGTCTCATTTCTTCAGCAGCTGTTGCCACACGAGCAATGTTTACACAAGCT 60
Db 9 ATGGCTTCCTGTCTCATTTCTTCAGCAGCTGTTGCCACACGAGCAATGTTTCTCAAGCT 68

Qy 61 AGCATGGTTGCACCTTTCAGTGGTCTCAAAATCTTCAGGCACCTTTCCCTGTTTACAAAGAAG 120
Db 69 AACATGGTTGCACCTTTCAGTGGCTTAAAGTGGCTTCAATCCCTGTTTCAAGGAAG 128

Qy 121 CAAAACCTTGACATCATTCCATTCCTAGCAATGTTGGAAGAGTTAGCTGCATGCAG 178
Db 129 CAAAACCTTGACATCATTCCATTCCTAGCAATGTTGGAAGAGTTAGCTGCATGCAG 186

RESULT 11
US-08-095-726-12
; Sequence 12, Application US/08095726
; Patent No. 5530188
; GENERAL INFORMATION:
; APPLICANT: Ausich, Rodney L
; APPLICANT: Brinkhaus, Friedhelm L
; APPLICANT: Mukharji, Indrani
; APPLICANT: Proffitt, John H
; APPLICANT: Yarger, James G
; APPLICANT: Yen, Huel-Che B
; TITLE OF INVENTION: Beta-Carotene Biosynthesis in

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/152,483B
; FILING DATE: No. 5529909ember 12, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 7/857,602
; FILING DATE: March 30, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 923,692
; FILING DATE: July 31, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 600,244
; FILING DATE: October 22, 1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 641,617
; FILING DATE: January 16, 1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 737,899
; FILING DATE: July 26, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P.
; REGISTRATION NUMBER: 25,227
; REFERENCE/DOCKET NUMBER: BIOG-20240/8129-040
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 854-3660
; TELEFAX: (415) 854-3694
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1442
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM:
; IMMEDIATE SOURCE:
; CLONE:
; FEATURE:
;
US-08-152-483B-8

Query Match 19.5%; Score 133.2; DB 2; Length 1442;
Best Local Similarity 84.3%; Pred. No. 5.9e-36;
Matches 150; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 1 ATGGCTTCCTGTCTCATTTCTTCAGCAGCTGTTGCCACACGAGCAATGTTTACACAAGCT 60
Db 9 ATGGCTTCCTGTCTCATTTCTTCAGCAGCTGTTGCCACACGAGCAATGTTTCTCAAGCT 68

Qy 61 AGCATGGTTGCACCTTTCAGTGGTCTCAAAATCTTCAGGCACCTTTCCCTGTTTACAAAGAAG 120
Db 69 AACATGGTTGCACCTTTCAGTGGCTTAAAGTGGCTTCAATCCCTGTTTCAAGGAAG 128

Qy 121 CAAAACCTTGACATCATTCCATTCCTAGCAATGTTGGAAGAGTTAGCTGCATGCAG 178
Db 129 CAAAACCTTGACATCATTCCATTCCTAGCAATGTTGGAAGAGTTAGCTGCATGCAG 186

RESULT 11
US-08-095-726-12
; Sequence 12, Application US/08095726
; Patent No. 5530188
; GENERAL INFORMATION:
; APPLICANT: Ausich, Rodney L
; APPLICANT: Brinkhaus, Friedhelm L
; APPLICANT: Mukharji, Indrani
; APPLICANT: Proffitt, John H
; APPLICANT: Yarger, James G
; APPLICANT: Yen, Huel-Che B
; TITLE OF INVENTION: Beta-Carotene Biosynthesis in

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/152,483B
; FILING DATE: No. 5529909ember 12, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 7/857,602
; FILING DATE: March 30, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 923,692
; FILING DATE: July 31, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 600,244
; FILING DATE: October 22, 1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 641,617
; FILING DATE: January 16, 1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 737,899
; FILING DATE: July 26, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P.
; REGISTRATION NUMBER: 25,227
; REFERENCE/DOCKET NUMBER: BIOG-20240/8129-040
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 854-3660
; TELEFAX: (415) 854-3694
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1442
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM:
; IMMEDIATE SOURCE:
; CLONE:
; FEATURE:
;
US-08-152-483B-8

Query Match 19.1%; Score 130.6; DB 2; Length 177;
Best Local Similarity 83.6%; Pred. No. 1.3e-35;
Matches 148; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy 1 ATGGCTTCCTGTCTCATTTCTTCAGCAGCTGTTGCCACACGAGCAATGTTTACACAAGCT 60
Db 1 ATGGCTTCCTGTCTCATTTCTTCAGCAGCTGTTGCCACACGAGCAATGTTTCTCAAGCT 60

Qy 61 AGCATGGTTGCACCTTTCAGTGGTCTCAAAATCTTCAGGCACCTTTCCCTGTTTACAAAGAAG 120
Db 61 AACATGGTTGCAGCTTTCAGTGGCTTAAAGTGGCTTCAATCCCTGTTTCAAGGAAG 120

Qy 121 CAAAACCTTGACATCATTCCATTCCTAGCAATGTTGGAAGAGTTAGCTGCATGCAG 177
Db 121 CAAAACCTTGACATCATTCCATTCCTAGCAATGTTGGAAGAGTTAGCTGCATGCAG 177

RESULT 12
US-08-096-043-12
; Sequence 12, Application US/08096043
; Patent No. 5530189
; GENERAL INFORMATION:
; APPLICANT: Ausich, Rodney L
; APPLICANT: Brinkhaus, Friedhelm L
; APPLICANT: Mukharji, Indrani
; APPLICANT: Proffitt, John H
; APPLICANT: Yarger, James G
; APPLICANT: Yen, Huel-Che B
; TITLE OF INVENTION: Lycopene Biosynthesis in
; TITLE OF INVENTION: Genetically Engineered Hosts
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amoco Corp., Patents and Licensing Dept
; STREET: 200 E Randolph St
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60680-0703
```

```
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.24
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/096,043
/ FILING DATE: 22-JUL-1993
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/785,568
/ FILING DATE: 30-OCT-1991
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Galloway, No. 5530189val B
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 3128567180
/ TELEFAX: 3128564972
/ INFORMATION FOR SEQ ID NO: 12:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 177 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ US-08-096-043-12

Query Match          19.1%; Score 130.6; DB 2; Length 177;
Best Local Similarity 83.6%; Pred. No. 1.3e-35;
Matches 148; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1 ATGGCTTCCTCTGTCATTTCTTCAGCAGCTGTGGCCACAGCGAGCAATGTTACACAAGCT 60
Db 1 ATGGCTTCCTCAGTCTCTTCCTCTGCAGCAGTGTGCCACCGCGAGCAATGTTGCTCAAGCT 60

QY 61 AGCATGGTTGCACCTTTTCACTGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 120
Db 61 AACATGGTGGCGCTTTTCACTGGCCTTAAAGTCAGCTGCCTCATTTCCCTGTTTCAAGGAAG 120

QY 121 CAAACCTTGACATCACTTCCATTGCTAGCAATGGTGGAGAGTTAGCTGCATGCAG 177
Db 121 CAAACCTTGACATCACTTCCATTGCTAGCAATGGTGGAGAGTTAGCTGCATGCAG 177

RESULT 13
US-093-577-8
/ Sequence 8, Application US/08093577
/ Patent No. 5545816
/ GENERAL INFORMATION:
/ APPLICANT: Ausich, Rodney L
/ APPLICANT: Brinkhaus, Friedhelm L
/ APPLICANT: Mukharji, Indrani
/ APPLICANT: Proffitt, John H
/ APPLICANT: Yarger, James G
/ APPLICANT: Yen, Huel-Che B
/ TITLE OF INVENTION: Phytoene Biosynthesis in
/ TITLE OF INVENTION: Genetically Engineered Hosts
/ NUMBER OF SEQUENCES: 43
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Amoco Corp., Patents and Licensing Dept
/ STREET: 200 E Randolph St
/ CITY: Chicago
/ STATE: IL
/ COUNTRY: USA
/ ZIP: 60680-0703
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.24
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/093,577
/ FILING DATE: 19-JUL-1993
/ CLASSIFICATION: 435
```

```
/
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/785,569
/ FILING DATE: 30-OCT-1991
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Galloway, No. 5545816val B
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 3128567180
/ TELEFAX: 3128564972
/ INFORMATION FOR SEQ ID NO: 8:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 177 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ US-08-093-577-8

Query Match          19.1%; Score 130.6; DB 2; Length 177;
Best Local Similarity 83.6%; Pred. No. 1.3e-35;
Matches 148; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1 ATGGCTTCCTCTGTCATTTCTTCAGCAGCTGTGGCCACAGCGAGCAATGTTACACAAGCT 60
Db 1 ATGGCTTCCTCAGTCTCTTCCTCTGCAGCAGTGTGCCACCGCGAGCAATGTTGCTCAAGCT 60

QY 61 AGCATGGTTGCACCTTTTCACTGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 120
Db 61 AACATGGTGGCGCTTTTCACTGGCCTTAAAGTCAGCTGCCTCATTTCCCTGTTTCAAGGAAG 120

QY 121 CAAACCTTGACATCACTTCCATTGCTAGCAATGGTGGAGAGTTAGCTGCATGCAG 177
Db 121 CAAACCTTGACATCACTTCCATTGCTAGCAATGGTGGAGAGTTAGCTGCATGCAG 177

RESULT 14
US-08-331-004A-6
/ Sequence 6, Application US/08331004A
/ Patent No. 5618988
/ GENERAL INFORMATION:
/ APPLICANT: Hauptmann, Randal
/ APPLICANT: Eschenfeldt, William H
/ APPLICANT: English, Jami
/ APPLICANT: Brinkhaus, Friedhelm L
/ TITLE OF INVENTION: Enhanced Carotenoid Accumulation
/ TITLE OF INVENTION: in Storage Organs of Genetically
/ TITLE OF INVENTION: Engineered Plants
/ NUMBER OF SEQUENCES: 9
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Amoco Corporation, Law Dept
/ STREET: 55 Shuman Boulevard, Suite 600
/ CITY: Naperville
/ STATE: IL
/ COUNTRY: USA
/ ZIP: 60563-8437
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.24
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/331,004A
/ FILING DATE:
/ CLASSIFICATION: 800
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Galloway, No. 5618988val B
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 7087172447
/ TELEFAX: 7087172430
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 177 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
```

```
;
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-331-004A-6

Query Match      19.1%; Score 130.6; DB 2; Length 177;
Best Local Similarity 83.6%; Pred. No. 1.3e-35;
Matches 148; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy 1 ATGGCTTCTCTGTCTATTTCTTCAGCAGCTGTGGCACAGCAGCAATGTTACACAAGCT 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;
Qy 61 AGCATGGTTGCACCTTTCACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;
Qy 61 AACATGGTGGCGCCCTTCACTGGCCTTAAAGTCAGCTGCTCATTCCTGTTTCAAGGAAG 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;
Qy 121 CAAAACCTTGACATCACTTCCATTGCTAGCAATGGTGGAAAGAGTTAGCTGCATGCAG 177
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;
Qy 121 CAAAACCTTGACATCACTTCCATTGCTAGCAATGGTGGAAAGAGTGCATGCAG 177
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
US-08-096-623A-12
; Sequence 12, Application US/08096623A
; Patent No. 5684238
; GENERAL INFORMATION:
; APPLICANT: Ausich, Rodney L.
; APPLICANT: Brinkhaus, Friedhelm L.
; APPLICANT: Mukharji, Indrani
; APPLICANT: Proffitt, John H.
; APPLICANT: Yarger, James G.
; APPLICANT: Yen, Huel-Che B.
; TITLE OF INVENTION: Biosynthesis of Zeaxanthin and
; TITLE OF INVENTION: Glycosylated Zeaxanthin in Genetically Engineered Hosts
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Welsh & Katz, Ltd.
; STREET: 120 S. Riverside Plaza, 22nd Floor
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/096,623A
; FILING DATE: 22-JUL-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/805,061
; FILING DATE: 09-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/662,921
; FILING DATE: 28-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/562,674
; FILING DATE: 03-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/525,551
; FILING DATE: 18-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/487,613
; FILING DATE: 02-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gamson, Edward P.
; REGISTRATION NUMBER: 29,381
; REFERENCE/DOCKET NUMBER: AMO-006.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 655-1500
; TELEFAX: (312) 655-1501
```

```
;
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 177 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; POSITION IN GENOME:
; MAP POSITION: 1 to 177
; UNITS: bp
US-08-096-623A-12

Query Match      19.1%; Score 130.6; DB 2; Length 177;
Best Local Similarity 83.6%; Pred. No. 1.3e-35;
Matches 148; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy 1 ATGGCTTCTCTGTCTATTTCTTCAGCAGCTGTGGCACAGCAGCAATGTTACACAAGCT 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;
Qy 61 AGCATGGTTGCACCTTTCACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;
Qy 61 AACATGGTGGCGCCCTTCACTGGCCTTAAAGTCAGCTGCTCATTCCTGTTTCAAGGAAG 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;
Qy 121 CAAAACCTTGACATCACTTCCATTGCTAGCAATGGTGGAAAGAGTTAGCTGCATGCAG 177
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;
Qy 121 CAAAACCTTGACATCACTTCCATTGCTAGCAATGGTGGAAAGAGTGCATGCAG 177
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: May 29, 2006, 12:19:38
Job time : 324 secs
```

18	144	21.1	736	10	US-10-487-901-5043	Sequence 5043, Ap
19	142.4	20.8	297	10	US-10-487-901-7209	Sequence 7209, Ap
20	142.4	20.8	377	10	US-10-487-901-3504	Sequence 3504, Ap
21	142.4	20.8	489	10	US-10-487-901-3503	Sequence 3503, Ap
22	142.4	20.8	504	10	US-10-487-901-7207	Sequence 7207, Ap
23	140.8	20.6	204	3	US-09-839-477-3	Sequence 3, Appli
24	140.8	20.6	204	8	US-10-758-064-3	Sequence 3, Appli
25	137.6	20.1	668	10	US-10-487-901-7208	Sequence 7208, Ap
26	129.2	18.9	174	3	US-09-854-286-15	Sequence 15, Appli
27	128.8	18.8	169	6	US-10-165-420-3	Sequence 3, Appli
28	124.6	18.2	683	10	US-10-487-901-1713	Sequence 1713, Ap
29	124.6	18.2	737	10	US-10-487-901-5045	Sequence 5045, Ap
30	124.6	18.2	737	10	US-10-487-901-5766	Sequence 5766, Ap
31	121.4	17.7	738	10	US-10-487-901-1716	Sequence 1716, Ap
32	121.4	17.7	753	10	US-10-487-901-5764	Sequence 5764, Ap
33	121.4	17.7	754	10	US-10-487-901-5028	Sequence 5028, Ap
34	119.8	17.5	714	10	US-10-487-901-1703	Sequence 1703, Ap
35	119.8	17.5	736	10	US-10-487-901-5044	Sequence 5044, Ap
36	115	16.8	704	9	US-10-425-115-85509	Sequence 85509, A
37	113	16.5	151	7	US-10-321-434-4	Sequence 4, Appli
38	111.8	16.3	654	10	US-10-487-901-5042	Sequence 5042, Ap
39	111	16.2	180	10	US-10-487-901-88	Sequence 88, Appli
40	97.2	14.2	149	3	US-09-839-477-5	Sequence 5, Appli
41	97.2	14.2	149	8	US-10-758-064-5	Sequence 5, Appli
42	85.2	12.5	363	3	US-09-770-791-468	Sequence 468, App
43	85.2	12.5	711	3	US-09-910-664-32	Sequence 32, Appli
44	85.2	12.5	711	8	US-10-333-184-31	Sequence 31, Appli
45	85.2	12.5	739	8	US-10-333-184-182	Sequence 182, App

## ALIGNMENTS

```

RESULT 1
US-09-855-341-7
; Sequence 7, Application US/09855341
; Patent No. US2002002715A1
; GENERAL INFORMATION:
; APPLICANT: VIITANEN, PAUL V.
; APPLICANT: MEYER, KNOT
; APPLICANT: VAN DYK, DREW
; TITLE OF INVENTION: HIGH LEVEL PRO
; TITLE OF INVENTION: IN GREEN PLAN
; FILE REFERENCE: BC1015 US NA
; CURRENT APPLICATION NUMBER: US/09/
; CURRENT FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 7
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of
US-09-855-341-7

```

```

181 TGGCATATGTACACCCCGGGTTAACGCAACTGCGTGGCTTATGTAAGAGATC 240
241 CTTCCCTTGGATCCGAACTGCTCGACTTGGCTGTTGCTGGAGGATTCATGACAAAACGT 300
241 CTTCCCTTGGATCCGAACTGCTCGACTTGGCTGTTGCTGGAGGATTCATGACAAAACGT 300
301 TTTGAACAGCAGGAGGAAAAACGGTAAAGCGTGACATGATCCGCGAAGGGTTTGTCCAGCAG 360
301 TTTGAACAGCAGGAGGAAAAACGGTAAAGCGTGACATGATCCGCGAAGGGTTTGTCCAGCAG 360
361 AATGAATATCCCGAAGAACTGCGCGCTGCTGCCCGAAGAGTCTCGTTACTGGTTACGTGAA 420
361 AATGAATATCCCGAAGAACTGCGCGCTGCTGCCCGAAGAGTCTCGTTACTGGTTACGTGAA 420
421 ATTTTGTATTATGTGCCGATGGTGAACCTGGCTTCCCGGTGCTACCGTCTTCTGTGCTCA 480
421 ATTTTGTATTATGTGCCGATGGTGAACCTGGCTTCCCGGTGCTACCGTCTTCTGTGCTCA 480
481 ACGTTAAGCGGGCCGGAGCTGGCGCTTACAAAAATTTGGGTAAAAACCGCGTTAGGACGCTAT 540
481 ACGTTAAGCGGGCCGGAGCTGGCGCTTACAAAAATTTGGGTAAAAACCGCGTTAGGACGCTAT 540
541 CTGTTTCATCATCATCGACATTAAACCCGGGACTTTATTGAGATAGGCCCTGATGCCGGCTG 600
541 CTGTTTCATCATCATCGACATTAAACCCGGGACTTTATTGAGATAGGCCCTGATGCCGGCTG 600
601 TGGGGGGCAGGTTCCCGCGCTGCGATTAAAGCGGTAAACCGCTTGTCTAAACAGACTGTTT 660
601 TGGGGGGCAGGTTCCCGCGCTGCGATTAAAGCGGTAAACCGCTTGTCTAAACAGACTGTTT 660
661 TTACCGGCGCTCACCGTTGTACTAA 684
661 TTACCGGCGCTCACCGTTGTACTAA 684

RESULT 2
US-09-896-866B-16
; Sequence 16 Application US/09896866B
; Patent No. US202015102A1
; GENERAL INFORMATION:
; APPLICANT: Flint, Dennis
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugates
; TITLE OF INVENTION: Benzoic Acid Glucosides
; FILE REFERENCE: BC1034 US NA
; CURRENT APPLICATION NUMBER: US/09/896,866B
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/216,615
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: open reading frame of the chloroplast-targeted CPL fusion protein
US-09-896-866B-16

```

QY	121	CAAAACCTTGACATCATTCCATTGCTAGCAATGGTGAAGAGTTAGCTGCATGCAGGTG	180
DB	121	CAAAACCTTGACATCATTCCATTGCTAGCAATGGTGAAGAGTTAGCTGCATGCAGGTG	180
QY	181	TGGCATATGTACACCCCGCGTTAAACGCAATGCGTGCCTGCGCTATTGTTAAAGAGATC	240
DB	181	TGGCATATGTACACCCCGCGTTAAACGCAATGCGTGCCTGCGCTATTGTTAAAGAGATC	240
QY	241	CCTGCCCTGGATTCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCCATGACAAAACGT	300
DB	241	CCTGCCCTGGATTCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCCATGACAAAACGT	300
QY	301	TTTGAACAGCAGGGAAGAAACCGGTAAAGCTGACGATGATCCCGGAAGGGTTTGTGAGCAG	360
DB	301	TTTGAACAGCAGGGAAGAAACCGGTAAAGCTGACGATGATCCCGGAAGGGTTTGTGAGCAG	360
QY	361	AATGAATATCCCGGAAGAACTCCCGCTGCTGCCGAAGAGTCTCGTTACTGGTTACGTGAA	420
DB	361	AATGAATATCCCGGAAGAACTCCCGCTGCTGCCGAAGAGTCTCGTTACTGGTTACGTGAA	420
QY	421	ATTTTGTATATGCGCGATGCTGAAACCGTGGCTTGCCTGCTGTAACCGCTTCTCTGTGTC	480
DB	421	ATTTTGTATATGCGCGATGCTGAAACCGTGGCTTGCCTGCTGTAACCGCTTCTCTGTGTC	480
QY	481	ACGTTAAGCGGGCGGAGCTGGCGTTACAAAATTTGGGTAAAAACCGCGTTAGGACGCTAT	540
DB	481	ACGTTAAGCGGGCGGAGCTGGCGTTACAAAATTTGGGTAAAAACCGCGTTAGGACGCTAT	540
QY	541	CTGTTACATCATCGACATTAAACCGGGACATTTATTGAGATAGGCGGTGATCCCGGGCTG	600
DB	541	CTGTTACATCATCGACATTAAACCGGGACATTTATTGAGATAGGCGGTGATCCCGGGCTG	600
QY	601	TGGGGGCGACGTTCCCGCTCGGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTT	660
DB	601	TGGGGGCGACGTTCCCGCTCGGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTT	660
QY	661	TTACCGGGGCTCACCGTTGTACTAA	684
DB	661	TTACCGGGGCTCACCGTTGTACTAA	684

RESULT 3

US-10-359-369-41

; Sequence 41, Application US/10359369

; Publication No. US20030215927A1

; GENERAL INFORMATION:

; APPLICANT: E. I. duPont de Nemours and Company, Inc.

; APPLICANT: Meyer, Knut

; APPLICANT: Van Dyk, Drew

; TITLE OF INVENTION: UDP-Glucosyltransferases

; FILE REFERENCE: CL1821 US NA

; CURRENT APPLICATION NUMBER: US/10/359,369

; CURRENT FILING DATE: 2003-02-06

; PRIOR APPLICATION NUMBER: 60/355,511

; PRIOR FILING DATE: 2002-02-07

; NUMBER OF SEQ ID NOS: 46

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 41

; LENGTH: 684

; TYPE: DNA

; ORGANISM: Escherichia coli

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(684)

US-10-359-369-41

Query Match	100.0%	Score 684;	DB 7;	Length 684;
Best Local Similarity	100.0%	Pred. No. 1.9e-222;		
Matches 684;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
Qv	1	ATGCGTTCTCTGTTCATTTCTTCAGCAGCTGTGCCACACGCAAGATGTTTACACAAGCT	60	

```
; OTHER INFORMATION: open reading frame of the chloroplast-targeted CPL fusion prote
us-10-699-050-16

Query Match          100.0%; Score 684; DB 8; Length 684;
Best Local Similarity 100.0%; Pred. No. 1.9e-222;
Matches 684; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCTTCCCTCTGTCAATTTCTTAGCAGCTGTTGCCACACGCAGCAATGTTACAAAGCT 60
Db | |||||
Qy 1 ATGCTTCCCTCTGTCAATTTCTTAGCAGCTGTTGCCACACGCAGCAATGTTACAAAGCT 60
Db | |||||

Qy 61 AGCATGGTTGCACCTTTTCACCTGGTCTCAAATCTTCAGCCAATTTCCTCTGTTACAAGAAG 120
Db | |||||
Qy 61 AGCATGGTTGCACCTTTTCACCTGGTCTCAAATCTTCAGCCAATTTCCTCTGTTACAAGAAG 120
Db | |||||

Qy 121 CAARAACTTGACATCACCTTCATTCGTAGCAATGGTGAAGAGTTAGCTGCATGCAGGGTG 180
Db | |||||
Qy 121 CAARAACTTGACATCACCTTCATTCGTAGCAATGGTGAAGAGTTAGCTGCATGCAGGGTG 180
Db | |||||

Qy 181 TGGCATATGTCACACCCCOCGCTTAACCGCAACTGGGTGCGCTGCGCTATTGTTAAAGAGATC 240
Db | |||||
Qy 181 TGGCATATGTCACACCCCOCGCTTAACCGCAACTGGGTGCGCTGCGCTATTGTTAAAGAGATC 240
Db | |||||

Qy 241 CTTGCCCTGGATTCGGAACCTGCTCGA CTGGCTGTTGCTGGAGGATTCATGACAAAAAGT 300
Db | |||||
Qy 241 CTTGCCCTGGATTCGGAACCTGCTCGA CTGGCTGTTGCTGGAGGATTCATGACAAAAAGT 300
Db | |||||

Qy 301 TTTCGAAACGACGAGGGAANAACGGTAAGCGTGACGATGATCCCGGAAGGCTTTGTCGAGCAG 360
Db | |||||
Qy 301 TTTCGAAACGACGAGGGAANAACGGTAAGCGTGACGATGATCCCGGAAGGCTTTGTCGAGCAG 360
Db | |||||

Qy 361 AATGAAATCCCCGAAGAACCTCCGCTGCTGCGGAAGAGTCTCGTTACTGTTACGTGAA 420
Db | |||||
Qy 361 AATGAAATCCCCGAAGAACCTCCGCTGCTGCGGAAGAGTCTCGTTACTGTTACGTGAA 420
Db | |||||

Qy 421 ATTGTTGTTATGTGCGCATGGTGAAACCGTGCGCTTTCGCGGTGCGTACCCTGCTTCCTGTCA 480
Db | |||||
Qy 421 ATTGTTGTTATGTGCGCATGGTGAAACCGTGCGCTTTCGCGGTGCGTACCCTGCTTCCTGTCA 480
Db | |||||

Qy 481 ACGTTAAGCGGGCGGAGCTGGCGTTACAAAATAATGGGTAAAAACCGCGTTAGGACGCTAT 540
Db | |||||
Qy 481 ACGTTAAGCGGGCGGAGCTGGCGTTACAAAATAATGGGTAAAAACCGCGTTAGGACGCTAT 540
Db | |||||

Qy 541 CTGTTACATCATCAGCAATTAACCCGGGACCTTTATTGAGATAGGCGGTGATCCGGGGCTG 600
Db | |||||
Qy 541 CTGTTACATCATCAGCAATTAACCCGGGACCTTTATTGAGATAGGCGGTGATCCGGGGCTG 600
Db | |||||

Qy 601 TGGGGGCGAGCTTCCCGCCTCGCATTAAGCGGTAAAAACCGCTGTTGCTAACGAACTGTTT 660
Db | |||||
Qy 601 TGGGGGCGAGCTTCCCGCCTCGCATTAAGCGGTAAAAACCGCTGTTGCTAACGAACTGTTT 660
Db | |||||

Qy 661 TTACCGGGGCTCACCGTTGTACTAA 684
Db | |||||
Qy 661 TTACCGGGGCTCACCGTTGTACTAA 684
Db | |||||
```

RESULT 5  
US-10-718-311-7  
; Sequence 7, Application US/10718311  
; Publication No. US20040143867A1  
; GENERAL INFORMATION:  
; APPLICANT: VIITANEN, PAUL V.  
; APPLICANT: MEYER, KNUT  
; APPLICANT: VAN DYK, DREW  
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID  
; TITLE OF INVENTION: IN GREEN PLANTS  
; FILE REFERENCE: BC1015 US NA  
; CURRENT APPLICATION NUMBER: US/10/718,311  
; CURRENT FILING DATE: 2003-11-20  
; PRIOR APPLICATION NUMBER: US/09/855,341  
; PRIOR FILING DATE: 2001-05-15  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: MICROSOFT OFFICE 97







```
; CURRENT APPLICATION NUMBER: US/10/462,162
; CURRENT FILING DATE: 2003-06-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 47
; LENGTH: 3452
; TYPE: DNA
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Nucleotide sequence of nucleic acid fragment encoding CPL (from
; OTHER INFORMATION: Escherichia coli), pHEA 1-hydroxylase (from Candida
; OTHER INFORMATION: parapsilosis), and UGT72B1 (from Arabidopsis thaliana).
US-10-462-162-47

Query Match          73.2%; Score 501; DB 9; Length 3452;
Best Local Similarity 100.0%; Pred. No. 1.8e-159;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 CATATGTCACACCCCGGTTAACGCAACTGCGTGCCTGCGCTATTGTAAGAGATCCCT 243
Db 1 CATATGTCACACCCCGGTTAACGCAACTGCGTGCCTGCGCTATTGTAAGAGATCCCT 60

QY 244 GCCTGGATCCGCAACTGCTCGACTGCGTGTGCTGGAGGATCCATGACAAAACGTTTT 303
Db 61 GCCTGGATCCGCAACTGCTCGACTGCGTGTGCTGGAGGATCCATGACAAAACGTTTT 120

QY 304 GAACAGCAGGAGAAAACGGTAAAGCGTGAAGCGTGAAGCGTGAAGCGTGAAGCGTGAAG 363
Db 121 GAACAGCAGGAGAAAACGGTAAAGCGTGAAGCGTGAAGCGTGAAGCGTGAAGCGTGAAG 180

QY 364 GAAATCCCGAAGAACTGCGGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATT 423
Db 181 GAAATCCCGAAGAACTGCGGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATT 240

QY 424 TTGTTATGTCGCGATGGTGAACCGTGGCTTGCCTGCGTACCGTCTCTCTGTGCAACG 483
Db 241 TTGTTATGTCGCGATGGTGAACCGTGGCTTGCCTGCGTACCGTCTCTCTGTGCAACG 300

QY 484 TTAAGCGGCGCGAGCTGCGGCTTACAAAAATTGGGTAAAAACGCGTTAGGACGCTATCTG 543
Db 301 TTAAGCGGCGCGAGCTGCGGCTTACAAAAATTGGGTAAAAACGCGTTAGGACGCTATCTG 360

QY 544 TTCACATCATCGACATTAACCCGGGACTTTATTGAGTAGGCGCGTGATGCCGGCTGTGG 603
Db 361 TTCACATCATCGACATTAACCCGGGACTTTATTGAGTAGGCGCGTGATGCCGGCTGTGG 420

QY 604 GGGCGACGTTCCCGCTCGATTAAGCGGTTAAACCGCTGTTGCTAACAGAACTGTTTTTA 663
Db 421 GGGCGACGTTCCCGCTCGATTAAGCGGTTAAACCGCTGTTGCTAACAGAACTGTTTTTA 480

QY 664 CCGGGCTCACCCGTTGTACTAA 684
Db 481 CCGGGCTCACCCGTTGTACTAA 501

RESULT 8
US-10-462-162-54
; Sequence 54, Application US/10462162
; Publication No. US20040261147A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; APPLICANT: Flint, Dennis
; TITLE OF INVENTION: High Level Production of Arbutin in Green Plants and Microbes
; FILE REFERENCE: CL 2155 US NA
; CURRENT APPLICATION NUMBER: US/10/462,162
; CURRENT FILING DATE: 2003-06-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 54
; LENGTH: 1971
; TYPE: DNA
```

```
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Nucleic acid sequence of a nucleic acid fragment inserted into an
; OTHER INFORMATION: expression vector pET29a encoding CPL (from Escherichia coli) an
; OTHER INFORMATION: pHEA 1-hydroxylase (from Candida parapsilosis).
US-10-462-162-54

Query Match          73.0%; Score 499.4; DB 9; Length 1971;
Best Local Similarity 99.8%; Pred. No. 4.6e-159;
Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 184 CATATGTCACACCCCGGTTAACGCAACTGCGTGCCTGCGCTATTGTAAGAGATCCCT 243
Db 1 CATATGTCACACCCCGGTTAACGCAACTGCGTGCCTGCGCTATTGTAAGAGATCCCT 60

QY 244 GCCTGGATCCGCAACTGCTCGACTGCGTGTGCTGGAGGATCCATGACAAAACGTTTT 303
Db 61 GCCTGGATCCGCAACTGCTCGACTGCGTGTGCTGGAGGATCCATGACAAAACGTTTT 120

QY 304 GAACAGCAGGAGAAAACGGTAAAGCGTGAAGCGTGAAGCGTGAAGCGTGAAGCGTGAAG 363
Db 121 GAACAGCAGGAGAAAACGGTAAAGCGTGAAGCGTGAAGCGTGAAGCGTGAAGCGTGAAG 180

QY 364 GAAATCCCGAAGAACTGCGGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATT 423
Db 181 GAAATCCCGAAGAACTGCGGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATT 240

QY 424 TTGTTATGTCGCGATGGTGAACCGTGGCTTGCCTGCGTACCGTCTCTCTGTGCAACG 483
Db 241 TTGTTATGTCGCGATGGTGAACCGTGGCTTGCCTGCGTACCGTCTCTCTGTGCAACG 300

QY 484 TTAAGCGGCGCGAGCTGCGGCTTACAAAAATTGGGTAAAAACGCGTTAGGACGCTATCTG 543
Db 301 TTAAGCGGCGCGAGCTGCGGCTTACAAAAATTGGGTAAAAACGCGTTAGGACGCTATCTG 360

QY 544 TTCACATCATCGACATTAACCCGGGACTTTATTGAGTAGGCGCGTGATGCCGGCTGTGG 603
Db 361 TTCACATCATCGACATTAACCCGGGACTTTATTGAGTAGGCGCGTGATGCCGGCTGTGG 420

QY 604 GGGCGACGTTCCCGCTCGATTAAGCGGTTAAACCGCTGTTGCTAACAGAACTGTTTTTA 663
Db 421 GGGCGACGTTCCCGCTCGATTAAGCGGTTAAACCGCTGTTGCTAACAGAACTGTTTTTA 480

QY 664 CCGGGCTCACCCGTTGTACTAA 684
Db 481 CCGGGCTCACCCGTTGTACTAA 501

RESULT 9
US-10-462-162-29
; Sequence 29, Application US/10462162
; Publication No. US20040261147A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; APPLICANT: Flint, Dennis
; TITLE OF INVENTION: High Level Production of Arbutin in Green Plants and Microbes
; FILE REFERENCE: CL 2155 US NA
; CURRENT APPLICATION NUMBER: US/10/462,162
; CURRENT FILING DATE: 2003-06-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 29
; LENGTH: 498
; TYPE: DNA
; ORGANISM: Escherichia coli
; ORGANISM: Escherichia coli
US-10-462-162-29

Query Match          72.8%; Score 498; DB 9; Length 498;
Best Local Similarity 100.0%; Pred. No. 6.9e-159;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



Db 241 TTATGTGCCGATGGTGAACCGTGGCTTGGCGGTGTAACCGTCTGTTCTCTGTCAACGGTTA 300  
QY 487 AGCGGCGCGGAGCTGGCGTTACAAAAATTGGGTAAACGCGTTAGGAGCTATCTGTTTC 546  
Db 301 AGCGGCGCGGAGCTGGCGTTACAAAAATTGGGTAAACGCGTTAGGAGCTATCTGTTTC 360  
QY 547 ACATCATCGACATTAAACCGGGACTTTATTGAGTAGGCGCGTGATCCGCGGCTGTGGGG 606  
Db 361 ACATCATCGACATTAAACCGGGACTTTATTGAGTAGGCGCGTGATCCGCGGCTGTGGGG 420  
QY 607 CGAGGTTCCCGCTCGGATTAAACCGGTAAACCGCTGTGCTAAACAGAACTGTTTTACCG 666  
Db 421 CGAGGTTCCCGCTCGGATTAAACCGGTAAACCGCTGTGCTAAACAGAACTGTTTTACCG 480  
QY 667 GCGTACCGTTGTAC 681  
Db 481 GCGTACCGTTGTAC 495

RESULT 12  
US-10-359-369-37  
; Sequence 37, Application US/10359369  
; Publication No. US20030215927A1  
; GENERAL INFORMATION:  
; APPLICANT: E. I. duPont de Nemours and Company, Inc.  
; APPLICANT: Meyer, Knut  
; APPLICANT: Viitanen, Paul  
; APPLICANT: Van Dyk, Drew  
; TITLE OF INVENTION: UDP-Glucosyltransferases  
; FILE REFERENCE: CL1821 US NA  
; CURRENT APPLICATION NUMBER: US/10/359,369  
; CURRENT FILING DATE: 2003-02-06  
; PRIOR APPLICATION NUMBER: 60/355,511  
; PRIOR FILING DATE: 2002-02-07  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 37  
; LENGTH: 495  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(495)  
US-10-359-369-37

Query Match 72.4%; Score 495; DB 7; Length 495;  
Best Local Similarity 100.0%; Pred. No. 7.3e-158;  
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 187 ATGTCAACCCCGCGTTAAACGCAACTGCGTGGCGTCTATTGTTAAAGAGATCCCTGCC 246  
Db 1 ATGTCAACCCCGCGTTAAACGCAACTGCGTGGCGTCTATTGTTAAAGAGATCCCTGCC 60  
QY 247 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTTGAA 306  
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTTGAA 120  
QY 307 CAGCAGGGAACCGTAAAGCTGACGATGATCCGCGAAGGGTTGTCGAGCAGATGAA 366  
Db 121 CAGCAGGGAACCGTAAAGCTGACGATGATCCGCGAAGGGTTGTCGAGCAGATGAA 180  
QY 367 ATCCCGGAAGAACTGCCGCTGTGCGAAAGAGTCTCGTTACTGTTAGTGAATTTTG 426  
Db 181 ATCCCGGAAGAACTGCCGCTGTGCGAAAGAGTCTCGTTACTGTTAGTGAATTTTG 240  
QY 427 TTATGTCCGATGGTGAACCGGTGCTGCGGTACCGTCTGTTCTCTGTCAACGGTTA 486  
Db 241 TTATGTCCGATGGTGAACCGGTGCTGCGGTACCGTCTGTTCTCTGTCAACGGTTA 300  
QY 487 AGCGGCGCGGAGCTGGCGTTACAAAAATTGGGTAAACCGCGTTAGGAGCTATCTGTTTC 546  
Db 301 AGCGGCGCGGAGCTGGCGTTACAAAAATTGGGTAAACCGCGTTAGGAGCTATCTGTTTC 360

QY 547 ACATCATCGACATTAAACCGGGACTTTATTGAGTAGGCGCGTGTGATCCGCGGCTGTGGGG 606  
Db 361 ACATCATCGACATTAAACCGGGACTTTATTGAGTAGGCGCGTGTGATCCGCGGCTGTGGGG 420  
QY 607 CGAGGTTCCCGCTCGGATTAAACCGGTAAACCGCTGTGCTAAACAGAACTGTTTTACCG 666  
Db 421 CGAGGTTCCCGCTCGGATTAAACCGGTAAACCGCTGTGCTAAACAGAACTGTTTTACCG 480  
QY 667 GCGTACCGTTGTAC 681  
Db 481 GCGTACCGTTGTAC 495  
RESULT 13  
US-10-699-050-12  
; Sequence 12, Application US/10699050  
; Publication No. US20040142437A1  
; GENERAL INFORMATION:  
; APPLICANT: Flint, Dennis  
; APPLICANT: Meyer, Knut  
; APPLICANT: Viitanen, Paul  
; TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugate  
; FILE REFERENCE: BC1034 US NA  
; CURRENT APPLICATION NUMBER: US/10/699,050  
; CURRENT FILING DATE: 2003-10-30  
; PRIOR APPLICATION NUMBER: US/09/896,866B  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: 60/216,615  
; PRIOR FILING DATE: 2000-07-07  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 12  
; LENGTH: 495  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; US-10-699-050-12

Query Match 72.4%; Score 495; DB 8; Length 495;  
Best Local Similarity 100.0%; Pred. No. 7.3e-158;  
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 187 ATGTCAACCCCGCGTTAAACGCAACTGCGTGGCGTCTATTGTTAAAGAGATCCCTGCC 246  
Db 1 ATGTCAACCCCGCGTTAAACGCAACTGCGTGGCGTCTATTGTTAAAGAGATCCCTGCC 60  
QY 247 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTTGAA 306  
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTTGAA 120  
QY 307 CAGCAGGGAACCGTAAAGCTGACGATGATCCGCGAAGGGTTGTCGAGCAGATGAA 366  
Db 121 CAGCAGGGAACCGTAAAGCTGACGATGATCCGCGAAGGGTTGTCGAGCAGATGAA 180  
QY 367 ATCCCGGAAGAACTGCCGCTGTGCGAAAGAGTCTCGTTACTGTTAGTGAATTTTG 426  
Db 181 ATCCCGGAAGAACTGCCGCTGTGCGAAAGAGTCTCGTTACTGTTAGTGAATTTTG 240  
QY 427 TTATGTCCGATGGTGAACCGGTGCTGCGGTACCGTCTGTTCTCTGTCAACGGTTA 486  
Db 241 TTATGTCCGATGGTGAACCGGTGCTGCGGTACCGTCTGTTCTCTGTCAACGGTTA 300  
QY 487 AGCGGCGCGGAGCTGGCGTTACAAAAATTGGGTAAACCGCGTTAGGAGCTATCTGTTTC 546  
Db 301 AGCGGCGCGGAGCTGGCGTTACAAAAATTGGGTAAACCGCGTTAGGAGCTATCTGTTTC 360  
QY 547 ACATCATCGACATTAAACCGGGACTTTATTGAGTAGGCGCGTGATCCGCGGCTGTGGGG 606  
Db 361 ACATCATCGACATTAAACCGGGACTTTATTGAGTAGGCGCGTGATCCGCGGCTGTGGGG 420  
QY 607 CGAGGTTCCCGCTCGGATTAAACCGGTAAACCGCTGTGCTAAACAGAACTGTTTTACCG 666  
Db 421 CGAGGTTCCCGCTCGGATTAAACCGGTAAACCGCTGTGCTAAACAGAACTGTTTTACCG 480

QY 667 GCGTCACCGTTGTAC 681  
Db 481 GCGTCACCGTTGTAC 495

RESULT 14

US-10-718-311-3  
; Sequence 3, Application US/10718311  
; Publication No. US20040143867A1  
; GENERAL INFORMATION:  
; APPLICANT: MEYER, KNUT  
; APPLICANT: VIITANEN, PAUL V.  
; APPLICANT: VAN DYK, DREW  
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID  
; FILE REFERENCE: IN GREEN PLANTS  
; CURRENT APPLICATION NUMBER: US/10/718,311  
; PRIOR FILING DATE: 2003-11-20  
; PRIOR APPLICATION NUMBER: US/09/855,341  
; PRIOR FILING DATE: 2001-05-15  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: MICROSOFT OFFICE 97  
; SEQ ID NO 3  
; LENGTH: 495  
; TYPE: DNA  
; ORGANISM: Unknown Organism  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism:E. coli  
US-10-718-311-3

Query Match 72.4%; Score 495; DB 8; Length 495;  
Best Local Similarity 100.0%; Pred. No. 7.3e-158;  
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 187 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATCCCTGCC 246  
Db 1 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATCCCTGCC 60  
QY 247 CTGGATCCGCAACTGCTGCACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGA 306  
Db 61 CTGGATCCGCAACTGCTGCACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGA 120  
QY 307 CAGCAGGGAACCGTAAGCGTACGATGATCCGCGAAGGTTTGTGCGAGCAATGAA 366  
Db 121 CAGCAGGGAACCGTAAGCGTACGATGATCCGCGAAGGTTTGTGCGAGCAATGAA 180  
QY 367 ATCCCGGAAGAACTGCCGCTGTCGCGAAGAGTCTCGTTACTGGTTACGTGAAATTTTG 426  
Db 181 ATCCCGGAAGAACTGCCGCTGTCGCGAAGAGTCTCGTTACTGGTTACGTGAAATTTTG 240  
QY 427 TTATGTCCGATGGTGAACCGTGGCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 486  
Db 241 TTATGTCCGATGGTGAACCGTGGCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
QY 487 AGCGGGCGGAGCTGGCGTTACAAAATTGGGTAAACCGCGTTAGGAGCGCTATCTGTC 546  
Db 301 AGCGGGCGGAGCTGGCGTTACAAAATTGGGTAAACCGCGTTAGGAGCGCTATCTGTC 360  
QY 547 ACATCATCGACATTAACCGGAGCTTTATTAGATAGGCGCGTGTGCGGGCTGTGGGG 606  
Db 361 ACATCATCGACATTAACCGGAGCTTTATTAGATAGGCGCGTGTGCGGGCTGTGGGG 420  
QY 607 CGACGTTCCCGCTCGGATTAAGCGGTAACCGCTGTTGCTTAACAGAACTGTTTTACCG 666  
Db 421 CGACGTTCCCGCTCGGATTAAGCGGTAACCGCTGTTGCTTAACAGAACTGTTTTACCG 480  
QY 667 GCGTCACCGTTGTAC 681  
Db 481 GCGTCACCGTTGTAC 495

RESULT 15

US-10-450-763-26139  
; Sequence 26139, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 790CIP3/US  
; CURRENT APPLICATION NUMBER: US/10/450,763  
; CURRENT FILING DATE: 2003-06-11  
; PRIOR APPLICATION NUMBER: PCT/US01/08631  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 26139  
; LENGTH: 1207  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIMILAR  
; LOCATION: (893)..(925)  
; OTHER INFORMATION: 81% homologous to Escherichia coli 4-hydroxybenzoate  
; OTHER INFORMATION: synthetase, accession number X57434, Smith-Waterman Score=49.  
US-10-450-763-26139

Query Match 69.1%; Score 472.8; DB 10; Length 1207;  
Best Local Similarity 99.2%; Pred. No. 4.5e-150;  
Matches 496; Conservative 0; Mismatches 2; Indels 2; Gaps 2;  
QY 187 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATCCCTGCC 246  
Db 652 ATGTCACACCCCGGTTAACGCAACTGCGTGGCTGCGCTATTGTAAGAGATCCCTGCC 711  
QY 247 CTGGATCCGCAACTGCTGCACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGA 306  
Db 712 CTGGATCCGCAACTGCTGCACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGA 771  
QY 307 CAGCAGGGAACCGTAAGCGTACGATGATCCGCGAAGGTTTGTGCGAGCAATGAA 366  
Db 772 CAGCAGGGAACCGTAAGCGTACGATGATCCGCGAAGGTTTGTGCGAGCAATGAA 831  
QY 367 ATCCCGGAAGAACTGCCGCTGTCGCGAAGAGTCTCGTTACTGGTTACGTGAAATTTTG 426  
Db 832 ATCCCGGAAGAACTGCCGCTGTCGCGAAGAGTCTCGTTACTGGTTACGTGAAATTTTG 891  
QY 427 T-TATGTCCGATGGTGAACCGTGGCTTGGCC- GGTCGTACCGTCTGTTCTGTCAAAGT 484  
Db 892 TGTATGTCCGATGGCGAACC CGGCTTGC CGGTCGTACCGTCTGTTCTGTCAAAGT 951  
QY 485 TAAGCGGCGGAGCTGGCGTTACAAAATTGGGTAAACCGCGTTAGGAGCGCTATCTGT 544  
Db 952 TAAGCGGCGGAGCTGGCGTTACAAAATTGGGTAAACCGCGTTAGGAGCGCTATCTGT 1011  
QY 545 TCACATCATCGACATTAACCGGAGCTTTATTAGATAGGCGCGTGTATCCGGGCTGTGGG 604  
Db 1012 TCACATCATCGACATTAACCGGAGCTTTATTAGATAGGCGCGTGTATCCGGGCTGTGGG 1071  
QY 605 GGCGAGCTTCCCGCTGGGATTAAGCGGTAACCGCTGTCGTAACAGAACTGTTTTTAC 664  
Db 1072 GGCGAGCTTCCCGCTGGGATTAAGCGGTAACCGCTGTCGTAACAGAACTGTTTTTAC 1131  
QY 665 CGGCGTCACCGTTGTACTAA 684  
Db 1132 CGGCGTCACCGTTGTACTAA 1151

Search completed: May 30, 2006, 01:43:38  
Job time : 1265 secs

November 2005

Published\_Applications\_Nucleic Acid and Published\_Applications\_Amino Acid database searches now generate two sets of results each. The Published\_Applications databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Newly published applications will appear in the Published\_Applications\_New databases; older published applications make up the Published\_Applications\_Main databases.

Searches run against Nucleic Acid Published\_Applications produce two sets of results, with the extensions **.rnpbm** (Published\_Applications\_NA\_Main) and **.rnpbn** (Published\_Applications\_NA\_New).

Searches run against Amino Acid Published\_Applications produce two sets of results, with the extensions **.rapbm** (Published\_Applications\_AA\_Main) and **.rapbn** (Published\_Applications\_AA\_New).



GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 30, 2006, 01:24:28 ; Search time 44 Seconds  
(without alignments)  
1754.839 Million cell updates/sec

Title: US-10-718-311-7  
Perfect score: 684  
Sequence: 1 atggcttcctctgtcatctt.....cggcgctcacgttgtactaa 684

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 242596 seqs, 56442199 residues

Total number of hits satisfying chosen parameters: 485192

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA New.\*

- 1: /EMC Celerra\_SIDS3/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 2: /EMC Celerra\_SIDS3/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 3: /EMC Celerra\_SIDS3/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 4: /EMC Celerra\_SIDS3/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 5: /EMC Celerra\_SIDS3/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 6: /EMC Celerra\_SIDS3/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 7: /EMC Celerra\_SIDS3/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*
- 8: /EMC Celerra\_SIDS3/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	85.2	12.5	718	6	US-10-953-349-10521
2	82	12.0	743	6	US-10-953-349-8894
3	74.4	10.9	799	6	US-10-953-349-9569
4	35.6	5.2	792	7	US-11-217-529-6134
5	32	4.7	2144	7	US-11-293-697-312
6	31.2	4.6	336	7	US-11-217-529-77187
7	30.8	4.5	2586	7	US-11-217-529-1832
8	30.4	4.4	807	6	US-10-953-349-40123
9	30	4.4	1801	6	US-10-953-349-7256
10	29.8	4.4	711	7	US-11-217-529-76273
11	29.8	4.4	3449	7	US-11-293-697-830
12	29.2	4.3	459	7	US-11-301-554-1043
13	29	4.2	1584	7	US-11-217-529-80201
14	28.8	4.2	450	7	US-11-301-554-1056
15	28.8	4.2	9399	7	US-11-236-836-2
16	28.8	4.2	11918	7	US-11-236-836-22
17	28.8	4.2	12023	7	US-11-236-836-20
18	28.8	4.2	13042	7	US-11-236-836-17
19	28.8	4.2	13042	7	US-11-236-836-23
20	28.8	4.2	13147	7	US-11-236-836-18
21	28.8	4.2	13147	7	US-11-236-836-25
22	28.6	4.2	901	6	US-10-953-349-1271
23	28.6	4.2	2160	7	US-11-293-697-438
24	28.4	4.2	855	7	US-11-217-529-77669
25	28.4	4.2	1107	7	US-11-217-529-78001

C	26	28.2	4.1	518	6	US-10-488-619-2571	Sequence 2571, Ap
C	27	28.2	4.1	1451	6	US-10-953-349-10879	Sequence 10879, A
C	28	28	4.1	2271	6	US-10-505-928-792	Sequence 792, App
C	29	28	4.1	3240	7	US-11-217-529-1227	Sequence 1227, Ap
C	30	27.6	4.0	1068	6	US-10-953-349-9573	Sequence 9573, Ap
C	31	27.6	4.0	3738	7	US-11-217-529-473	Sequence 473, Appl
C	32	27.6	4.0	3953	6	US-10-473-173-56	Sequence 56, Appl
C	33	27.4	4.0	1435	6	US-10-953-349-13607	Sequence 13607, A
C	34	27.4	4.0	1714	6	US-10-953-349-32909	Sequence 32909, A
C	35	27.2	4.0	726	6	US-10-953-349-40081	Sequence 40081, A
C	36	27.2	4.0	1320	7	US-11-311-555-5	Sequence 5, Appli
C	37	27.2	4.0	1320	7	US-11-311-561-5	Sequence 5, Appli
C	38	27.2	4.0	1320	7	US-11-101-316-155	Sequence 155, App
C	39	27.2	4.0	2568	7	US-11-217-529-75586	Sequence 75586, A
C	40	27	3.9	612	7	US-11-217-529-174132	Sequence 174132, A
C	41	27	3.9	1157	6	US-10-953-349-35133	Sequence 35133, A
C	42	27	3.9	1760	6	US-10-953-349-11108	Sequence 11108, A
C	43	27	3.9	138941	6	US-10-489-730-10	GENERAL INFORMATI
C	44	26.8	3.9	549	6	US-10-953-349-11985	Sequence 11985, A
C	45	26.8	3.9	774	7	US-11-217-529-174376	Sequence 174376,

ALIGNMENTS

RESULT 1

US-10-953-349-10521  
; Sequence 10521, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; TITLE OF INVENTION: ENCODED THERBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: Patentin version 3.3  
; SEQ ID NO 10521  
; LENGTH: 718  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-10-953-349-10521

Query Match 12.5%; Score 85.2; DB 6; Length 718;  
Best Local Similarity 70.7%; Pred. No. 1e-19;  
Matches 130; Conservative 0; Mismatches 48; Indels 6; Gaps 1;

QY	1	ATGGCTTCCTCTGTCTATTTCTTCAGCAGCTGTTCACACGACGACCAATGTTACACAGCT	60
DB	32	ATGGCTTCCTCTATGCTCTCTCTGCGCTGTGTTAC-----CTCCCCGGCTCAAGCC	85
QY	61	AGCATGGTGCACCTTTCACCTGCTCTCAATCTTCAGCCACTTTCCTGTTACAAAGAAG	120
DB	86	ACCATGGTCGCTCCATTCACCTGCTGTTGAAGTCACTCGCTCTCTTCGCGGTCAACCCGCAAG	145
QY	121	CAAAACCTTGACATCACTTCCATTGCTAGCAATGGTGGAGAGATGCTCATCAGGTG	180
DB	146	GCCAAACAGACATTAATCTTCCATCAACAAGCAATGGGGGAAGATTAGTGCATGAAGGTG	205
QY	181	TGGC 184	
DB	206	TGGC 209	

RESULT 2

US-10-953-349-8894  
; Sequence 8894, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; TITLE OF INVENTION: ENCODED THERBY





```

, , TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
, , FILE REFERENCE: S-38-285
, , CURRENT APPLICATION NUMBER: US/11/217,529
, , CURRENT FILING DATE: 2005-09-02
, , PRIOR APPLICATION NUMBER: US 10/932,182
, , PRIOR FILING DATE: 2004-09-02
, , NUMBER OF SEQ ID NOS: 197023
, , SOFTWARE: PatentIn version 3.3
, , SEQ ID NO 77187
, , LENGTH: 336
, , TYPE: DNA
, , ORGANISM: Saccharomyces pastorianus
US-11-217-529-77187

```

```

; SEQ ID NO 40123
; LENGTH: 807
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
US-10-953-349-40123

```

Qy	219	GCTGCGCTATTGTAAA	234
Dβ	156	GCCATGTTATCTGAAA	141

Query Match 4.4%; Score 30; DB 6; Length 1801;  
Best Local Similarity 48.8%; Pred. No. 1.6;  
Matches 81, Conservative 0; Mismatches 85; Indels 0; Gaps 0

Qy 6 TTCTCTGTGTCATTTCTTCAGCAGCTGTGGCACACGACGACATGTTACACAGCTAGCAT 65

Db 20 TTCCTTGTTCAACCTCCGAGAGATTACGATAGTGAAGAGAGCTACAAAATTTCCCAA 79  
QY 66 GGTTCACCTTTCTACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAACAAA 125  
Db 80 TTTATCAAAAAATAGAAATTTCTTACTTTTTTCAAAATTCATTTCTCCCAAAAAAATA 139  
QY 126 CTTGACATCACTTCCTCATGCTAGCAATGGTGAAGAGATTAGCTGC 171  
Db 140 TCTTGTGAAACTGCAACTGTGATCCATGGCGGAGGATATGGCTGC 185

## RESULT 10

US-11-217-529-76273  
; Sequence 76273, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHIISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR APPLICATION NUMBER: US 10/932,182  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 76273  
; LENGTH: 711  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-217-529-76273

Query Match 4.4%; Score 29.8; DB 7; Length 711;  
Best Local Similarity 49.1%; Pred. No. 1.1;  
Matches 79; Conservative 0; Mismatches 82; Indels 0; Gaps 0;  
QY 5 CTTCTCTGTCTATTCTTCAGCAGCTGTTGCCACAGCAAGCAATGTTACAAAGCTAGCA 64  
Db 488 CTTCTGAGCTAAGTCTTCTCTGCTGCTCCAGCTCCACTGAAGCTAAGCACTTCTG 547  
QY 65 TGGTTGACCTTTCTACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAACAAA 124  
Db 548 CTGCTCCAAGCTCCACTGCTGTGCCAAGACCTCTGCCATCTCTCAAAATTACCGATGTCAAA 607  
QY 125 ACCTTGACATCACTTCCATTGCTAGCAATGGTGGAGAGTT 165  
Db 608 TCCAAAGCTACCAAGGCTGTTTCTGAGCAAACTGAAAAACGGT 648

## RESULT 11

US-11-293-697-830/c  
; Sequence 830, Application US/11293697  
; Publication No. US20060105376A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/11/293,697  
; CURRENT FILING DATE: 2005-12-05  
; PRIOR APPLICATION NUMBER: US/10/108,260  
; PRIOR FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 830  
; LENGTH: 3449  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-293-697-830

Query Match 4.4%; Score 29.8; DB 7; Length 3449;  
Best Local Similarity 54.0%; Pred. No. 2.7;  
Matches 61; Conservative 0; Mismatches 52; Indels 0; Gaps 0;  
QY 80 CTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAACAAAACCTTTGACATCATT 139  
Db 1379 CTTCTTTAACGTTATCACACATACACATACAAAGGAGCAGCAGTACTACTTACT 1320  
QY 140 CCATTGCTAGCAATGGTGGAAAGAGTTAGCTGCATGCAGGTGTGGCATATGTCA 192  
Db 1319 CAACTGCAATGAAAAATAAAGTAAGATGCCATCAAAACTTTGAATAAGTCA 1267

## RESULT 12

US-11-301-554-1043  
; Sequence 1043, Application US/11301554  
; Publication No. US20060088527A1  
; GENERAL INFORMATION:  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Wang, Tongcong  
; APPLICANT: Watanabe, Yoshihiro  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: Johnson, Jeffrey C.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Durnam, Margarita  
; APPLICANT: Carter, Darrick  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Vedwick, Thomas S.  
; APPLICANT: Bangur, Chaitanya S.  
; APPLICANT: McHabb, Andria  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.478C21  
; CURRENT APPLICATION NUMBER: US/11/301,554  
; CURRENT FILING DATE: 2005-12-13  
; PRIOR APPLICATION NUMBER: US 10/283,017  
; PRIOR FILING DATE: 2002-10-28  
; PRIOR APPLICATION NUMBER: US 10/113,872  
; PRIOR FILING DATE: 2002-03-28  
; PRIOR APPLICATION NUMBER: US 10/017,754  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 09/902,941  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: US 09/849,626  
; PRIOR FILING DATE: 2001-05-03  
; PRIOR APPLICATION NUMBER: US 09/736,457  
; PRIOR FILING DATE: 2000-12-13  
; PRIOR APPLICATION NUMBER: US 09/702,705  
; PRIOR FILING DATE: 2000-10-30  
; PRIOR APPLICATION NUMBER: US 09/677,419  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 09/671,325  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US 09/658,824  
; PRIOR FILING DATE: 2000-09-08  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 2157  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1043  
; LENGTH: 459  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-301-554-1043

Query Match 4.3%; Score 29.2; DB 7; Length 459;  
Best Local Similarity 50.7%; Pred. No. 1.3;  
Matches 70; Conservative 0; Mismatches 68; Indels 0; Gaps 0;  
QY 253 CCGCAACTGCTCGACTGCTTCTGCTGGAGGATTCATGACAAAACGTTTGAACAGCAG 312  
Db 96 CAGGACCTGCTGGTTTCCCTGCTGTCGACAGCAATGGTGAACCTCGTGTGAAGGAG 155

QY 313 GGAAAAACGGTAAGCGTGACGATGATCCGCGAAGGGTTTCTCGAGCAGAAATGAAATCCCC 372  
Db 156 AAAGAGGGGCTCCGGGTGAGAAAGGTGAAGGAGGCCCTCTCGAGTTGCAGGACCCCTG 215  
QY 373 GAAGAACTGCCGTGCTG 390  
Db 216 GAGGTCTGACCTGCTG 233

## RESULT 13

US-11-217-529-80201  
; Sequence 80201, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHIISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; PRIOR FILING DATE: 2005-09-02  
; PRIOR APPLICATION NUMBER: US 10/932,182  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 80201  
; LENGTH: 1584  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-217-529-80201

Query Match 4.2%; Score 29; DB 7; Length 1584;

Best Local Similarity 51.1%; Pred. No. 3.2; Mismatches 0; Indels 0; Gaps 0;  
Matches 68; Conservative 0

QY 289 ATGACAAAAAGTTTGAACAGCAGGAGGAAAAACGGTAAGCGTGACGATGATCCGCGAAGGG 348  
Db 1 ATGCGGAGATTATGACACTAGCCATTAAAGCCACTAATACGGTGATGACATGGG 60  
QY 349 TTGTGTCGACAGAAATGAAATCCCCGAAGAACTGCCGCTGCTGCCGAAGAGTCTCGTTAC 408  
Db 61 TTGAAGAGCAATACCAAGTGACCAACTTTTACATCACTCGCGAATGATTTCATATTC 120  
QY 409 TGGTTACGTGAAA 421  
Db 121 TACTTGTGATGATA 133

## RESULT 14

US-11-301-554-1056/c  
; Sequence 1056, Application US/11301554  
; Publication No. US20060088527A1  
; GENERAL INFORMATION:  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Wang, Tonglong  
; APPLICANT: Watanabe, Yoshihiro  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: Johnson, Jeffrey C.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Durham, Margarita  
; APPLICANT: Carter, Darrick  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Bangur, Chaitanya S.  
; APPLICANT: McNabb, Andria  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.478C21  
; CURRENT APPLICATION NUMBER: US/11/301,554

; CURRENT FILING DATE: 2005-12-13  
; PRIOR APPLICATION NUMBER: US 10/283,017  
; PRIOR FILING DATE: 2002-10-28  
; PRIOR APPLICATION NUMBER: US 10/113,872  
; PRIOR FILING DATE: 2002-03-28  
; PRIOR APPLICATION NUMBER: US 10/017,754  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 09/902,941  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: US 09/849,626  
; PRIOR FILING DATE: 2001-05-03  
; PRIOR APPLICATION NUMBER: US 09/736,457  
; PRIOR FILING DATE: 2000-12-13  
; PRIOR APPLICATION NUMBER: US 09/702,705  
; PRIOR FILING DATE: 2000-10-30  
; PRIOR APPLICATION NUMBER: US 09/677,419  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 09/671,325  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US 09/658,824  
; PRIOR FILING DATE: 2000-09-08  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 2157  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1056  
; LENGTH: 450  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 21, 22, 230, 232, 377, 391  
; OTHER INFORMATION: n = A,T,C or G  
US-11-301-554-1056

Query Match 4.2%; Score 28.8; DB 7; Length 450;

Best Local Similarity 50.0%; Pred. No. 1.8;  
Matches 69; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 253 CGCGAACTCTCGACTGGCTGTTGCTGGAGGATTCATGACAAACGTTTGAACAGCAG 312  
Db 364 CAGGACCTCTGTTTCCCTGGTCTCTGGACAGAAATGTTGAACTGGTGGTAAAGGAG 305  
QY 313 GGAAAAACGGTAAGCGTGACGATGATCCGCGAAGGGTTTGTTCGAGCAGAAATGAAATCCCC 372  
Db 304 AAAGAGGGGCTCCGGTGAAGAGGTGAAGGAGGCCCTCTCTGAGTTGCAGGACCCCTG 245  
QY 373 GAAGAACTGCCGTGCTG 390  
Db 244 GAGTTCTGGACNTCTG 227

## RESULT 15

US-11-236-836-2/c  
; Sequence 2, Application US/11236836  
; Publication No. US20060105365A1  
; GENERAL INFORMATION:  
; APPLICANT: MARTIN, Annette  
; APPLICANT: GHIBAUDO, David  
; APPLICANT: COHEN, Lisette  
; APPLICANT: LEMON, STANLEY M.  
; TITLE OF INVENTION: Chimeric GB Virus B (GBV-B)  
; FILE REFERENCE: UTSG:270US  
; CURRENT APPLICATION NUMBER: US/11/236,836  
; CURRENT FILING DATE: 2005-09-27  
; PRIOR APPLICATION NUMBER: 60/613,266  
; PRIOR FILING DATE: 2004-09-27  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 9399  
; TYPE: DNA  
; ORGANISM: GBV-A-like virus  
US-11-236-836-2

Query Match 4.2%; Score 28.8; DB 7; Length 9399;  
Best Local Similarity 69.6%; Pred. No. 10;  
Matches 39; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 113 CAAGAAGCAAAACCTTGACATCACTTCCATTGCTAGCAATGGTGGAGAGTTAGC 168  
|||||  
Db 2720 CAAGAAGCAAAACGAGACATGACAACTAATGCTACTACAGGTGGTATTGTCAGC 2665  
|||||

Search completed: May 30, 2006, 04:14:20  
Job time : 44 secs

GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 30, 2006, 01:04:24 ; Search time 4485 Seconds  
(without alignments)  
8528.166 Million cell updates/sec

Title: US-10-718-311-7  
Perfect score: 684  
Sequence: 1 atgggttcctctgtcatttc.....cggcgccaccgtgtactaa 684

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 48236798 seqs, 27959665780 residues  
Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST.\*

- 1: gb\_est1.\*
- 2: gb\_est3.\*
- 3: gb\_est4.\*
- 4: gb\_est5.\*
- 5: gb\_est6.\*
- 6: gb\_hic1.\*
- 7: gb\_est2.\*
- 8: gb\_est7.\*
- 9: gb\_est8.\*
- 10: gb\_est9.\*
- 11: gb\_ges1.\*
- 12: gb\_ges2.\*
- 13: gb\_ges3.\*
- 14: gb\_ges4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	363	53.1	832	13	CL666350	CL666350 PRI0152b
2	295.4	43.2	828	13	CL687382	CL687382 PRI0146c
3	184	26.9	205	7	BE462847	BE462847 EST325226
4	184	26.9	235	7	AW039464	AW039464 EST281745
5	184	26.9	237	7	AW039041	AW039041 EST281014
6	184	26.9	238	7	AW041024	AW041024 EST283888
7	184	26.9	250	7	AW442951	AW442951 EST307881
8	184	26.9	260	7	AW037699	AW037699 EST279328
9	184	26.9	260	7	AW094122	AW094122 EST27302
10	184	26.9	267	1	AI773928	AI773928 EST255028
11	184	26.9	277	1	AI773502	AI773502 EST254602
12	184	26.9	279	2	BI931188	BI931188 EST551077
13	184	26.9	279	7	BE462486	BE462486 EST324752
14	184	26.9	283	7	AW094091	AW094091 EST27271
15	184	26.9	286	7	BF051382	BF051382 EST436557
16	184	26.9	288	7	AW037811	AW037811 EST279440
17	184	26.9	294	7	AW040243	AW040243 EST282749
18	184	26.9	329	7	AW038470	AW038470 EST280153
19	184	26.9	330	7	AW092013	AW092013 EST285193

RESULT 1  
CL666350/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITILE  
JOURNAL  
PUBMED  
COMMENT

ALIGNMENTS

CL666350 832 bp DNA linear GSS 09-JUL-2004  
PRI0152b.A03 - PRI0152b.B21 (832) Mixed stage fosmid library of P.  
pacificus var. California Pristionchus pacificus genomic, genomic  
survey sequence.

CL666350  
GSS  
Pristionchus pacificus

Pristionchus pacificus  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;

Neodiplogasteridae; Pristionchus.  
1 (bases 1 to 832)

Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.  
AppADB: an AcedB database for the nematode satellite organism

Pristionchus pacificus  
Nucleic Acids Res. 32 (1), D421-D422 (2004)

14681447

Contact: Sommer RJ

Evolutionary Biology

Max-Planck-Institute for Developmental Biology

Spemannstr. 37-39, Tuebingen D-72076, Germany

Tel: 00497071601371

Fax: 00497071601498

Email: ralf.sommer@tuebingen.mpg.de

This library was generated at Caltech, Pasadena, USA and end

sequenced at Vancouver, Canada.

Seq primer: T7

Class: fosmid ends.

Location/Qualifiers

1..832

/organism="Pristionchus pacificus"

/mol\_type="genomic DNA"

/strain="California"

/db\_xref="taxon:54126"

/clone\_lib="Mixed stage fosmid library of P. pacificus

var. California"

/note="Vector: pEpifos-5 Fosmid vector"

ORIGIN

Query Match

53.1%; Score 363; DB 13; Length 832;



Db 8 ATGGCTTCCTCTGTCAATTTCTTTCAGCAGCTGTTGCCACACGCGAGCAATGTTACACAAGCT 67  
 QY 61 AGCATGGTTCACCTTTCACCTGCTCAAAATCTTCAGCCATTTCCCTGTTACAAGAAG 120  
 Db 68 AGCATGGTTCACCTTTCACCTGCTCAAAATCTTCAGCCATTTCCCTGTTACAAGAAG 127  
 QY 121 CAAAACCTTGACATCACTTCCATTTAGCAATGGTGGGAAGAGTTAGCTGCATGCAGGTG 180  
 Db 128 CAAAACCTTGACATCACTTCCATTTAGCAATGGTGGGAAGAGTTAGCTGCATGCAGGTG 187  
 QY 181 TGGC 184  
 Db 188 TGGC 191

RESULT 4  
 AW039464  
 LOCUS  
 DEFINITION EST281745 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA  
 ACCESSION AW039464  
 VERSION AW039464.1 GI:5898218  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Lycopersicon esculentum (Solanum lycopersicum)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.  
 1 (bases 1 to 235)  
 D'Ascenzo, M., He, X., Lyman, J., Holt, I. E., Liang, F., Upton, J.,  
 Ronning, C. M., Craven, M. B., Fujii, C. Y., Bowman, C. L., Nierman, W.,  
 Fraser, C. M., Venter, J. C., Martin, G. B., Tanksley, S. D. and  
 Giovannoni, J.  
 Generation of ESTs from tomato leaf tissue  
 Unpublished (1999)  
 Contact: CUGI  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Email: <http://www.genome.clemson.edu/orders/index.html>  
 5 prime sequence.

## FEATURES

Location/Qualifiers  
 1..235  
 /organism="Lycopersicon esculentum"  
 /mol\_type="mRNA"  
 /cultivar="Rio Grande PtoR"  
 /db\_xref="taxon:4081"  
 /clone="cLET10112"  
 /tissue\_type="leaf"  
 /dev\_stage="4-6 week old plants"  
 /lab\_host="XL1-Blue MRF"  
 /clone\_lib="tomato mixed elicitor, BTI"  
 /note="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:  
 XhoI; cLET - Inoculated with a variety of disease response  
 elicitors. Plants exposed to 2,6 dichloroisonicotinic  
 acid, BTH, jasmonic acid, ethylene, fenthion, EIX,  
 okadaic acid, or systemin prior to tissue harvest. EcoRI  
 site was destroyed during cloning."

## ORIGIN

Query Match 26.9%; Score 184; DB 7; Length 235;  
 Best Local Similarity 100.0%; Pred. No. 3e-47;  
 Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGGCTTCCTCTGTCAATTTCTTTCAGCAGCTGTTGCCACACGCGAGCAATGTTACACAAGCT 60  
 Db 14 ATGGCTTCCTCTGTCAATTTCTTTCAGCAGCTGTTGCCACACGCGAGCAATGTTACACAAGCT 73  
 QY 61 AGCATGGTTCACCTTTCACCTGCTCAAAATCTTCAGCCATTTCCCTGTTACAAGAAG 120  
 Db 74 AGCATGGTTCACCTTTCACCTGCTCAAAATCTTCAGCCATTTCCCTGTTACAAGAAG 133

QY 121 CAAAACCTTGACATCACTTCCATTTAGCAATGGTGGGAAGAGTTAGCTGCATGCAGGTG 180  
 Db 134 CAAAACCTTGACATCACTTCCATTTAGCAATGGTGGGAAGAGTTAGCTGCATGCAGGTG 193  
 QY 181 TGGC 184  
 Db 194 TGGC 197

## RESULT 5

AW039041  
 LOCUS  
 DEFINITION EST281014 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA  
 ACCESSION AW039041  
 VERSION AW039041.1 GI:5897795  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Lycopersicon esculentum (Solanum lycopersicum)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.  
 1 (bases 1 to 237)  
 D'Ascenzo, M., He, X., Lyman, J., Holt, I. E., Liang, F., Upton, J.,  
 Ronning, C. M., Craven, M. B., Fujii, C. Y., Bowman, C. L., Nierman, W.,  
 Fraser, C. M., Venter, J. C., Martin, G. B., Tanksley, S. D. and  
 Giovannoni, J.  
 Generation of ESTs from tomato leaf tissue  
 Unpublished (1999)  
 Contact: CUGI  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Email: <http://www.genome.clemson.edu/orders/index.html>  
 5 prime sequence.

## FEATURES

Location/Qualifiers  
 1..237  
 /organism="Lycopersicon esculentum"  
 /mol\_type="mRNA"  
 /cultivar="Rio Grande PtoR"  
 /db\_xref="taxon:4081"  
 /clone="cLET10C2"  
 /tissue\_type="leaf"  
 /dev\_stage="4-6 week old plants"  
 /lab\_host="XL1-Blue MRF"  
 /clone\_lib="tomato mixed elicitor, BTI"  
 /note="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:  
 XhoI; cLET - Inoculated with a variety of disease response  
 elicitors. Plants exposed to 2,6 dichloroisonicotinic  
 acid, BTH, jasmonic acid, ethylene, fenthion, EIX,  
 okadaic acid, or systemin prior to tissue harvest. EcoRI  
 site was destroyed during cloning."

## ORIGIN

Query Match 26.9%; Score 184; DB 7; Length 237;  
 Best Local Similarity 100.0%; Pred. No. 3e-47;  
 Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGGCTTCCTCTGTCAATTTCTTTCAGCAGCTGTTGCCACACGCGAGCAATGTTACACAAGCT 60  
 Db 9 ATGGCTTCCTCTGTCAATTTCTTTCAGCAGCTGTTGCCACACGCGAGCAATGTTACACAAGCT 68  
 QY 61 AGCATGGTTCACCTTTCACCTGCTCAAAATCTTCAGCCATTTCCCTGTTACAAGAAG 120  
 Db 69 AGCATGGTTCACCTTTCACCTGCTCAAAATCTTCAGCCATTTCCCTGTTACAAGAAG 128  
 QY 121 CAAAACCTTGACATCACTTCCATTTAGCAATGGTGGGAAGAGTTAGCTGCATGCAGGTG 180  
 Db 129 CAAAACCTTGACATCACTTCCATTTAGCAATGGTGGGAAGAGTTAGCTGCATGCAGGTG 188  
 QY 181 TGGC 184  
 Db 189 TGGC 192

```

RESULT 6
AW041024
LOCUS
DEFINITION
  EST283888 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA
  clone cLET615, mRNA sequence.
ACCESSION
AW041024
VERSION
AW041024.1
KEYWORDS
GI:5899778
SOURCE
  Lycopersicon esculentum (Solanum lycopersicum)
ORGANISM
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
  asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE
1 (bases 1 to 238)
AUTHORS
  D'Ascenzo,M., He,X., Lyman,J., Holt,I.E., Liang,F., Upton,J.,
  Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W.,
  Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and
  Giovannoni,J.
  Generation of ESTs from tomato leaf tissue
  Unpublished (1999)
  Contact: CUGI
  Clemson University Genomics Institute
  Clemson University
  100 Jordan Hall, Clemson, SC 29634, USA
  Email: http://www.genome.clemson.edu/orders/index.html
  5 prime sequence.
FEATURES
  source
  1..238
    /organism="Lycopersicon esculentum"
    /mol_type="mRNA"
    /cultivar="Rio Grande Ptor"
    /db_xref="taxon:4081"
    /clone="cLET615"
    /tissue_type="leaf"
    /dev_stage="4-6 week old plants"
    /lab_host="XLI-Blue MRF"
    /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
    XhoI; cLET - Inoculated with a variety of disease response
    elicitors. Plants exposed to 2,6 dichloroisonicotinic
    acid, BTH, jasmonic acid, ethylene, fenthion, BIX,
    okadaic acid, or systemin prior to tissue harvest. EcoRI
    site was destroyed during cloning."
ORIGIN
  Query Match      26.9%; Score 184; DB 7; Length 238;
  Best Local Similarity 100.0%; Pred. No. 3e-47;
  Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCCTGTCATTTCTTTCAGCAGCTGTTGCCACGAGCAATGTTACACAAGCT 60
Db 19 ATGGCTTCCTGTCATTTCTTTCAGCAGCTGTTGCCACGAGCAATGTTACACAAGCT 78

QY 61 AGCATGGTGGACCTTCACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAGAAG 120
Db 79 AGCATGGTGGACCTTCACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAGAAG 138

QY 121 CAAAACCTTGACATCACTTCCATTGCTAGCAATGTTGGAAGAGTTAGCTGCATGCAGGTG 180
Db 139 CAAAACCTTGACATCACTTCCATTGCTAGCAATGTTGGAAGAGTTAGCTGCATGCAGGTG 198

QY 181 TGGC 184
Db 199 TGGC 202

RESULT 7
AW442951
LOCUS
DEFINITION
  EST307881 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA
  clone cLET42H11 5', mRNA sequence.
ACCESSION
AW442951
VERSION
AW442951.1
KEYWORDS
GI:5895133
SOURCE
  Lycopersicon esculentum (Solanum lycopersicum)
ORGANISM
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
  asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE
1 (bases 1 to 250)
AUTHORS
  D'Ascenzo,M., He,X., Lyman,J., Matern,A.L., Vision,T., Holt,I.E.,
  Liang,F., Hansen,T.S., Ronning,C.M., Craven,M.B., Bowman,C.L.,
  Nierman,W., Fraser,C.M., Venter,J.C., Tanksley,S.D.,
  Giovannoni,J.J. and Martin,G.B.
  Generation of ESTs from tomato callus (mixed elicitor)
  Unpublished (1999)
  Contact: CUGI
  Clemson University Genomics Institute
  Clemson University
  100 Jordan Hall, Clemson, SC 29634, USA
  Email: http://www.genome.clemson.edu/orders/index.html
  5 prime sequence.
FEATURES
  source
  1..250
    /organism="Lycopersicon esculentum"
    /mol_type="mRNA"
    /cultivar="Rio Grande Ptor"
    /db_xref="taxon:4081"
    /clone="cLET42H11"
    /tissue_type="leaf"
    /dev_stage="4-6 week old plants"
    /lab_host="XLI-Blue MRF"
    /clone_lib="tomato mixed elicitor, BTI"
    /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
    XhoI; cLET - Inoculated with a variety of disease response
    elicitors. Plants exposed to 2,6 dichloroisonicotinic
    acid, BTH, jasmonic acid, ethylene, fenthion, BIX,
    okadaic acid, or systemin prior to tissue harvest. EcoRI
    site was destroyed during cloning."
ORIGIN
  Query Match      26.9%; Score 184; DB 7; Length 250;
  Best Local Similarity 100.0%; Pred. No. 3.1e-47;
  Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCCTGTCATTTCTTTCAGCAGCTGTTGCCACGAGCAATGTTACACAAGCT 60
Db 6 ATGGCTTCCTGTCATTTCTTTCAGCAGCTGTTGCCACGAGCAATGTTACACAAGCT 65

QY 61 AGCATGGTGGACCTTCACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAGAAG 120
Db 66 AGCATGGTGGACCTTCACTGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAGAAG 125

QY 121 CAAAACCTTGACATCACTTCCATTGCTAGCAATGTTGGAAGAGTTAGCTGCATGCAGGTG 180
Db 126 CAAAACCTTGACATCACTTCCATTGCTAGCAATGTTGGAAGAGTTAGCTGCATGCAGGTG 185

QY 181 TGGC 184
Db 186 TGGC 189

RESULT 8
AW037699
LOCUS
DEFINITION
  EST279328 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA
  clone cLET4M17, mRNA sequence.
ACCESSION
AW037699
VERSION
AW037699.1
KEYWORDS
GI:5896453
SOURCE
  Lycopersicon esculentum (Solanum lycopersicum)
ORGANISM
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
  asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.

```



asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.  
1 (bases 1 to 260)  
D'Ascenzo, M., He, X., Lyman, J., Holt, I.E., Liang, F., Upton, J.,  
Ronning, C.M., Craven, M.B., Fujii, C.Y., Bowman, C.L., Nierman, W.,  
Fraser, C.M., Venter, J.C., Martin, G.B., Tanksley, S.D. and  
Giovannoni, J.

TITLE  
JOURNAL  
COMMENT  
Generation of ESTs from tomato leaf tissue  
Unpublished (1999)  
Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.

FEATURES  
source  
1..260  
/organism="Lycopersicon esculentum"  
/mol\_type="mRNA"  
/cultivar="Rio Grande PtoR"  
/db\_xref="taxon:4081"  
/clone="cLET4M17"  
/tissue\_type="leaf"  
/dev\_stage="4-6 week old plants"  
/lab\_host="XLI-Blue MRF"  
/clone\_lib="Tomato mixed elicitor, BTI"  
/notes="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:  
XhoI; cLET - Inoculated with a variety of disease response  
elicitors. Plants exposed to 2,6 dichloroisocotinic  
acid, BTH, jasmonic acid, ethylene, fenthion, EIX,  
okadaic acid, or systemin prior to tissue harvest. EcoRI  
site was destroyed during cloning."

## ORIGIN

Query Match 26.9%; Score 184; DB 7; Length 260;  
Best Local Similarity 100.0%; Pred. No. 3.1e-47;  
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCCTCTGTCATTTCTTCAGCAGCTGTTGCCACACGCGCAATGTTACACAAGCT 60  
|||||  
Db 12 ATGGCTTCCTCTGTCATTTCTTCAGCAGCTGTTGCCACACGCGCAATGTTACACAAGCT 71  
|||||  
QY 61 AGCATGGTGGACCTTTCACCTGCTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 120  
|||||  
Db 72 AGCATGGTGGACCTTTCACCTGCTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 131  
|||||  
QY 121 CAAACCTTGACATCACTTCCATTCAGCAATGGTGGAGAGTTAGCTGCATGCAGGTG 180  
|||||  
Db 132 CAAACCTTGACATCACTTCCATTCAGCAATGGTGGAGAGTTAGCTGCATGCAGGTG 191  
|||||

QY 181 TGGC 184  
|||||

Db 192 TGGC 195

RESULT 9  
AW094122  
LOCUS  
DEFINITION  
EST287302 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA  
clone cLET27G20, mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Lycopersicon esculentum (Solanum lycopersicum)  
Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE  
1 (bases 1 to 260)  
D'Ascenzo, M., He, X., Lyman, J., Holt, I.E., Liang, F., Upton, J.,  
Ronning, C.M., Craven, M.B., Fujii, C.Y., Bowman, C.L., Nierman, W.,  
Fraser, C.M., Venter, J.C., Martin, G.B., Tanksley, S.D. and  
Giovannoni, J.  
Generation of ESTs from tomato leaf tissue

## TITLE

JOURNAL  
COMMENT

Unpublished (1999)  
Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.

FEATURES  
source

1..260  
/organism="Lycopersicon esculentum"  
/mol\_type="mRNA"  
/cultivar="Rio Grande PtoR"  
/db\_xref="taxon:4081"  
/clone="cLET27G20"  
/tissue\_type="leaf"  
/dev\_stage="4-6 week old plants"  
/lab\_host="XLI-Blue MRF"  
/clone\_lib="tomato mixed elicitor, BTI"  
/notes="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:  
XhoI; cLET - Inoculated with a variety of disease response  
elicitors. Plants exposed to 2,6 dichloroisocotinic  
acid, BTH, jasmonic acid, ethylene, fenthion, EIX,  
okadaic acid, or systemin prior to tissue harvest. EcoRI  
site was destroyed during cloning."

## ORIGIN

Query Match 26.9%; Score 184; DB 7; Length 260;  
Best Local Similarity 100.0%; Pred. No. 3.1e-47;  
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCCTCTGTCATTTCTTCAGCAGCTGTTGCCACACGCGCAATGTTACACAAGCT 60  
|||||  
Db 6 ATGGCTTCCTCTGTCATTTCTTCAGCAGCTGTTGCCACACGCGCAATGTTACACAAGCT 65  
|||||  
QY 61 AGCATGGTGGACCTTTCACCTGCTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 120  
|||||  
Db 66 AGCATGGTGGACCTTTCACCTGCTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 125  
|||||  
QY 121 CAAACCTTGACATCACTTCCATTCAGCAATGGTGGAGAGTTAGCTGCATGCAGGTG 180  
|||||  
Db 126 CAAACCTTGACATCACTTCCATTCAGCAATGGTGGAGAGTTAGCTGCATGCAGGTG 185  
|||||

QY 181 TGGC 184  
|||||

Db 186 TGGC 189

RESULT 10  
AI773928

LOCUS  
DEFINITION  
EST255028 tomato resistant, Cornell Lycopersicon esculentum cDNA  
clone cLER8L7, mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Lycopersicon esculentum (Solanum lycopersicum)  
Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE  
1 (bases 1 to 267)

AUTHORS  
D'Ascenzo, M., He, X., Lyman, J., Matern, A.L., Vision, T., Holt, I.E.,  
Liang, F., Upton, J., Ronning, C.M., Craven, M.B., Fujii, C.Y.,  
Bowman, C.L., Nierman, W., Fraser, C.M., Venter, J.C., Tanksley, S.D.,  
Giovannoni, J. and Martin, G.B.  
Generation of ESTs from Pseudomonas resistant tomato

TITLE  
JOURNAL  
COMMENT

Unpublished (1999)  
Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.

FEATURES  
source

```

Location/Qualifiers
1..267
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="R11-12 (35S::Pto in Rio Grande x Money Maker)"
/db_xref="taxon:4081"
/clone="cLER8L7"
/tissue_type="leaf"
/dev_stage="4-week old"
/lab_host="SOLR"
/clone_lib="tomato resistant, Cornell"
/notes="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLER - Tomato Pseudomonas Resistant EST Library.
Directionally cloned cDNAs inserted into pBlueScript
SK(-) at 5' end with EcoRI and 3' end with XhoI site."

ORIGIN
Query Match      26.9%; Score 184; DB 1; Length 267;
Best Local Similarity 100.0%; Pred. No. 3.2e-47;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCCTGTCATTTCTCAGCAGCTGTTGCCACACGAGCAATGTTACACAAGCT 60
Db 23 ATGGCTTCCTGTCATTTCTCAGCAGCTGTTGCCACACGAGCAATGTTACACAAGCT 82

QY 61 AGCATGGTTGCACCTTTCACCTGGTCTCAAAATCTTCAGCCACTTTCCTGTTTACAAAGAAG 120
Db 83 AGCATGGTTGCACCTTTCACCTGGTCTCAAAATCTTCAGCCACTTTCCTGTTTACAAAGAAG 142

QY 121 CAAACCTTGACATCACTTCCATTCCTAGCAATGTTGGAAGAGTTAGCTGCATGCAGGTG 180
Db 143 CAAACCTTGACATCACTTCCATTCCTAGCAATGTTGGAAGAGTTAGCTGCATGCAGGTG 202

QY 181 TGGC 184
Db 203 TGGC 206

RESULT 11
AI773502
LOCUS      277 bp      mRNA      linear      EST 18-MAY-2001
DEFINITION EST254602 tomato resistant, Cornell Lycopersicon esculentum cDNA
clone cLER7A10, mRNA sequence.
ACCESSION AI773502
VERSION   AI773502
KEYWORDS  EST.
SOURCE    Lycopersicon esculentum (Solanum lycopersicum)
ORGANISM  Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 277)
D'Ascenzo,M., He,X., Lyman,J., Matern,A.L., Vision,T., Holt,I.E.,
Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y.,
Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Tanksley,S.D.,
Giovannoni,J.J. and Martin,G.B.
Generation of ESTs from Pseudomonas resistant tomato
Unpublished (1999)
Contact: CUGI
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
Location/Qualifiers
1..277
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="R11-12 (35S::Pto in Rio Grande x Money Maker)"
/db_xref="taxon:4081"
/clone="cLER7A10"
/tissue_type="leaf"
/dev_stage="4-week old"

FEATURES
source
Location/Qualifiers
1..279
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cTOC19C6"
/tissue_type="flower"
/dev_stage="buds 8mm to preanthesis"
/clone_lib="tomato flower, 8 mm to preanthesis buds"
/notes="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: pBlueScript SK(-); Site_1: EcoRI; Site_2:
Institute for Genomic Research; Flower buds and flowers
were taken from greenhouse plants (4-8 wks old, TA496).
They were immediately frozen in liquid nitrogen and then
size-separated while remaining frozen."

```

```

/lab_host="SOLR"
/clone_lib="tomato resistant, Cornell"
/notes="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLER - Tomato Pseudomonas Resistant EST Library.
Directionally cloned cDNAs inserted into pBlueScript
SK(-) at 5' end with EcoRI and 3' end with XhoI site."

ORIGIN
Query Match      26.9%; Score 184; DB 1; Length 277;
Best Local Similarity 100.0%; Pred. No. 3.2e-47;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCCTGTCATTTCTCAGCAGCTGTTGCCACACGAGCAATGTTACACAAGCT 60
Db 5 ATGGCTTCCTGTCATTTCTCAGCAGCTGTTGCCACACGAGCAATGTTACACAAGCT 64

QY 61 AGCATGGTTGCACCTTTCACCTGGTCTCAAAATCTTCAGCCACTTTCCTGTTTACAAAGAAG 120
Db 65 AGCATGGTTGCACCTTTCACCTGGTCTCAAAATCTTCAGCCACTTTCCTGTTTACAAAGAAG 124

QY 121 CAAACCTTGACATCACTTCCATTCCTAGCAATGTTGGAAGAGTTAGCTGCATGCAGGTG 180
Db 125 CAAACCTTGACATCACTTCCATTCCTAGCAATGTTGGAAGAGTTAGCTGCATGCAGGTG 184

QY 181 TGGC 184
Db 185 TGGC 188

RESULT 12
BI931188
LOCUS      279 bp      mRNA      linear      EST 18-OCT-2001
DEFINITION EST551077 tomato flower, 8 mm to preanthesis buds Lycopersicon
esculentum cDNA clone cTOC19C6 5' end, mRNA sequence.
ACCESSION BI931188
VERSION   BI931188
KEYWORDS  EST.
SOURCE    Lycopersicon esculentum (Solanum lycopersicum)
ORGANISM  Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 279)
van der Hoeven,R.S., Bezzerides,J.L., Karanycheva,S.A., Tsai,J.,
Utterback,T., Van Aken,S., Ronning,C.M., Nierman,W., Fraser,C.M.,
Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
Generation of ESTs from tomato flower tissue, buds 8 mm -
preanthesis
Unpublished (2001)
Contact: CUGI
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
This clone is available through the Clemson University Genomics
Institute
Seq primer: T3.
Location/Qualifiers
1..279
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cTOC19C6"
/tissue_type="flower"
/dev_stage="buds 8mm to preanthesis"
/clone_lib="tomato flower, 8 mm to preanthesis buds"
/notes="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: pBlueScript SK(-); Site_1: EcoRI; Site_2:
Institute for Genomic Research; Flower buds and flowers
were taken from greenhouse plants (4-8 wks old, TA496).
They were immediately frozen in liquid nitrogen and then
size-separated while remaining frozen."

```

## ORIGIN

Query Match 26.9%; Score 184; DB 2; Length 279;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-47;  
 Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCCTGTCATTTCTTCAGCAGCTGTTGCCACACGCGCAAGTTCACAAAGCT 60  
 |||||  
 Db 7 ATGGCTTCCTGTCATTTCTTCAGCAGCTGTTGCCACACGCGCAAGTTCACAAAGCT 66

QY 61 AGCATGGTTCACCTTTCTCACTGCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 120  
 |||||  
 Db 67 AGCATGGTTCACCTTTCTCACTGCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 126

QY 121 CAAACCTTGACATCACTTCCTATTTAGCAATGGTGGAAAGATTAGTCGATGCAGGTG 180  
 |||||  
 Db 127 CAAACCTTGACATCACTTCCTATTTAGCAATGGTGGAAAGATTAGTCGATGCAGGTG 186

QY 181 TGGC 184  
 |||||  
 Db 187 TGGC 190

## RESULT 13

BE462486 279 bp mRNA linear EST 18-MAY-2001  
 EST324752 tomato flower buds 0-3 mm, Cornell University  
 Lycopersicon esculentum cDNA clone CTOA1301, mRNA sequence.

ACCESSION BE462486  
 VERSION BE462486  
 KEYWORDS EST.

SOURCE Lycopersicon esculentum (Solanum lycopersicum)

ORGANISM Lycopersicon esculentum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE 1 (bases 1 to 279)  
 AUTHORS van der Hoeven,R.S., Bezzerides,J.L., Matern,A.L., Holt,I.E.,  
 Liang,F., Hansen,T., Craven,M.B., Bowman,C.L., Ronning,C.M.,  
 Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and  
 Tanksley,S.D.

TITLE Generation of ESTs from tomato flower tissue, 0-3 mm buds

JOURNAL Unpublished (1999)

COMMENT Contact: CUGI

Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Email: <http://www.genome.clemson.edu/orders/index.html>  
 5 prime sequence.

## FEATURES

Location/Qualifiers  
 1..279  
 /organism="Lycopersicon esculentum"  
 /mol\_type="mRNA"  
 /cultivar="TA496"  
 /db\_xref="taxon:4081"  
 /clone="CTOA1301"  
 /tissue\_type="flower"  
 /dev\_stage="0-3mm buds"  
 /clone\_lib="tomato flower buds 0-3 mm, Cornell University"  
 /notes="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
 XhoI; supplier: Tanksley; Flower buds and flowers were  
 taken from greenhouse plants (4-8 wks old, TA496). They  
 were immediately frozen in liquid nitrogen and then  
 size-separated while remaining frozen."

## ORIGIN

Query Match 26.9%; Score 184; DB 7; Length 279;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-47;  
 Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCCTGTCATTTCTTCAGCAGCTGTTGCCACACGCGCAAGTTCACAAAGCT 60  
 |||||  
 Db 7 ATGGCTTCCTGTCATTTCTTCAGCAGCTGTTGCCACACGCGCAAGTTCACAAAGCT 66

QY 61 AGCATGGTTCACCTTTCTCACTGCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 120  
 |||||  
 Db 67 AGCATGGTTCACCTTTCTCACTGCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 126

QY 121 CAAACCTTGACATCACTTCCTATTTAGCAATGGTGGAAAGATTAGTCGATGCAGGTG 180  
 |||||  
 Db 127 CAAACCTTGACATCACTTCCTATTTAGCAATGGTGGAAAGATTAGTCGATGCAGGTG 186

QY 181 TGGC 184  
 |||||  
 Db 187 TGGC 190

## RESULT 14

AW094091 283 bp mRNA linear EST 18-MAY-2001  
 EST287271 tomato mixed elicitor, B71 Lycopersicon esculentum cDNA  
 clone cLET27A10, mRNA sequence.

ACCESSION AW094091  
 VERSION AW094091  
 KEYWORDS EST.

SOURCE Lycopersicon esculentum (Solanum lycopersicum)

ORGANISM Lycopersicon esculentum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE 1 (bases 1 to 283)  
 AUTHORS D'Ascenzo,M., He,X., Lyman,J., Holt,I.E., Liang,F., Upton,J.,  
 Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W.,  
 Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and  
 Giovannoni,J.

TITLE Generation of ESTs from tomato leaf tissue

JOURNAL Unpublished (1999)

COMMENT Contact: CUGI

Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Email: <http://www.genome.clemson.edu/orders/index.html>  
 5 prime sequence.

## FEATURES

Location/Qualifiers  
 1..283  
 /organism="Lycopersicon esculentum"  
 /mol\_type="mRNA"  
 /cultivar="Rio Grande Ptor"  
 /db\_xref="taxon:4081"  
 /clone="cLET27A10"  
 /tissue\_type="leaf"  
 /dev\_stage="4-6 week old plants"  
 /lab\_host="X11-Blue MRF"  
 /clone\_lib="tomato mixed elicitor, BTI"  
 /notes="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
 XhoI; cLEF - Inoculated with a variety of disease response  
 elicitors. Plants exposed to 2,6 dichloroisonicotinic  
 acid, BTH, jasmonic acid, ethylene, fenthion, Etx,  
 okadaic acid, or systemin prior to tissue harvest. EcoRI  
 site was destroyed during cloning."

## ORIGIN

Query Match 26.9%; Score 184; DB 7; Length 283;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-47;  
 Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCCTGTCATTTCTTCAGCAGCTGTTGCCACACGCGCAAGTTCACAAAGCT 60  
 |||||  
 Db 12 ATGGCTTCCTGTCATTTCTTCAGCAGCTGTTGCCACACGCGCAAGTTCACAAAGCT 71

QY 61 AGCATGGTTCACCTTTCTCACTGCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 120  
 |||||  
 Db 72 AGCATGGTTCACCTTTCTCACTGCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 131

QY 121 CAAACCTTGACATCACTTCCTATTTAGCAATGGTGGAAAGATTAGTCGATGCAGGTG 180  
 |||||

```

Db      132 CAAACCTTGACATCATTCCATTGCTAGCAATGGTGGAAAGATTAGTGCATGCGAGGTG 191
Qy      181 TGGC 184
      ||||
Db      192 TGGC 195

RESULT 15
BF051382
LOCUS      286 bp      mRNA      linear      EST 18-MAY-2001
DEFINITION EST436557 tomato developing/immature green fruit Lycopersicon
            esculentum cDNA clone CLEM22P19 5' sequence, mRNA sequence.
ACCESSION  BF051382
VERSION    BF051382.1  GI:10805278
KEYWORDS
SOURCE
ORGANISM  Lycopersicon esculentum (Solanum lycopersicum)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE  1 (bases 1 to 286)
AUTHORS   Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
            Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
            Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and
            Tanksley,S.D.
TITLE     Generation of ESTs from tomato fruit tissue, immature green
JOURNAL   Unpublished (2000)
COMMENT   Contact: CUGI
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
            Location/Qualifiers
            1..286
            /organism="Lycopersicon esculentum"
            /mol_type="mRNA"
            /cultivar="TA496"
            /db_xref="taxon:4081"
            /clone="CLEM22P19"
            /tissue_type="fruit"
            /dev_stage="immature green (5-35 days post-anthesis)"
            /lab_host="SOLR"
            /clone_lib="tomato developing/immature green fruit"
            /note="Vector: pBluescriptSKmCtadapt; Site1: EcoRI;
            Site 2: XhoI; Fruit were tagged at 5 dpa (0.5 cm) and
            harvested at 7 day intervals through 35 dpa. Equal masses
            of tissue from each stage were combined (including seeds
            and locules) prior to mRNA isolation."

ORIGIN
Query Match      26.9%; Score 184; DB 7; Length 286;
Best Local Similarity 100.0%; Pred. No. 3.3e-47;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  ATGGCTTCCTCTGTCATTCTTCACGAGCTGTTCACACGACGCAATGTACACAAGCT 60
      |||||||
Db      10  ATGGCTTCCTCTGTCATTCTTCACGAGCTGTTCACACGACGCAATGTACACAAGCT 69

Qy      61  AGCATGGTTGCACCTTCACATGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 120
      |||||||
Db      70  AGCATGGTTGCACCTTCACATGGTCTCAAAATCTTCAGCCACTTTCCCTGTTACAAAGAAG 129

Qy      121 CAAACCTTGACATCATTCCATTGCTAGCAATGGTGGAAAGTTAGCTGCATGCGAGGTG 180
      |||||||
Db      130 CAAACCTTGACATCATTCCATTGCTAGCAATGGTGGAAAGTTAGCTGCATGCGAGGTG 189

Qy      181 TGGC 184
      ||||
Db      190 TGGC 193

```

Search completed: May 30, 2006, 03:27:40  
Job time : 4488 secs

GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 30, 2006, 01:30:02 / Search time 5152.27 Seconds

(without alignments)  
3164.934 Million cell updates/sec

Title: US-10-718-311-16

Perfect score: 887

Sequence: 1 MQVHMSHPALQTLRALRYC.....RLSGKPLLLTELFLPASPLY 170

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 6366136 seqs, 31973710525 residues

Total number of hits satisfying chosen parameters: 12732272

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\*\_p2n.model -DRV=xlh  
-Q=/abs/ABSSWEB\_spool/US10718311/runat\_26052006\_164916\_13439/app\_query.fasta\_1  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rgc -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=bl6sum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abses03h  
-USER=US10718311@CGN\_1\_5767@runat\_26052006\_164916\_13439 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*

- 1: gb\_env.\*
- 2: gb\_pat.\*
- 3: gb\_ph.\*
- 4: gb\_pl.\*
- 5: gb\_pr.\*
- 6: gb\_ro.\*
- 7: gb\_sts.\*
- 8: gb\_sy.\*
- 9: gb\_un.\*
- 10: gb\_vi.\*
- 11: gb\_ov.\*
- 12: gb\_hgs.\*
- 13: gb\_in.\*
- 14: gb\_om.\*
- 15: gb\_ba.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	887	100.0	684	2	AR428736 Sequence
2	887	100.0	684	2	AR455412 Sequence
3	887	100.0	684	2	AX329368 Sequence

4	887	100.0	684	2	AX382258	AX382258 Sequence	
5	854	96.3	495	2	A47932	A47932 Sequence 1	
6	854	96.3	495	2	AR428733	AR428733 Sequence	
7	854	96.3	495	2	AR455409	AR455409 Sequence	
8	854	96.3	495	2	AX329364	AX329364 Sequence	
9	854	96.3	495	2	AX382254	AX382254 Sequence	
10	854	96.3	854	2	AR144788	AR144788 Sequence	
11	854	96.3	1498	15	ECUBIFA	M96268 Escherichia	
12	854	96.3	2000	2	E11274	E11274 ubiC and ubiA	
13	854	96.3	2000	15	ECUBICA	M93136 Escherichia	
14	854	96.3	2000	15	ECUBIPLS	M93413 Escherichia	
15	854	96.3	2034	15	ECUBIAC	X57434 E.coli ubiC	
16	854	96.3	2348	15	ECUBI	X66619 E.coli gene	
17	854	96.3	6641	2	AR144789	AR144789 Sequence	
18	854	96.3	110000	15	U00096_42	Continuation (43 o	
19	854	96.3	110000	15	AP009048_42	Continuation (43 o	
20	854	96.3	176195	15	ECOUW89	U00006 E. coli chr	
21	850	95.8	110000	15	AE005174_51	Continuation (52 o	
22	846	95.4	110000	15	BA000007_50	Continuation (51 o	
23	846	95.4	110000	15	BA000007_51	Continuation (52 o	
24	843	95.0	1396	15	DQ087228	DQ087228 Escherich	
c	25	843	95.0	110000	15	AE005674_43	Continuation (44 o
	26	843	95.0	110000	15	AE014073_34	Continuation (35 o
c	27	843	95.0	110000	15	CP000036_41	Continuation (42 o
	28	839	94.6	110000	15	CP000034_42	Continuation (43 o
c	29	835	94.1	110000	15	CP000038_44	Continuation (45 o
	30	833	93.9	110000	15	AE014075_47	Continuation (48 o
31	798	90.0	1918	2	E07849	E07849 gDNA encodi	
32	725	81.7	23880	15	AE008898	AE008898 Salmonell	
33	725	81.7	110000	15	CP000026_41	Continuation (42 o	
34	725	81.7	110000	15	CP000026_42	Continuation (43 o	
35	721	81.3	110000	15	AE014613_42	Continuation (43 o	
36	721	81.3	110000	15	AE017220_43	Continuation (44 o	
37	721	81.3	234050	15	AL627282	AL627282 Salmonell	
38	673	75.9	645	2	AR384505	AR384505 Sequence	
39	437.5	49.3	10668	15	AE013658	AE013658 Versinia	
40	437.5	49.3	110000	15	EX936398_04	Continuation (5 of	
41	437.5	49.3	214050	15	AJ414142	AJ414142 Versinia	
42	437.5	49.3	290803	15	AE017128	AE017128 Versinia	
43	424.5	47.9	110000	15	EX950851_06	Continuation (7 of	
c	44	417	47.0	349107	15	EX571873	EX571873 Photobab
	45	408.5	46.1	110000	15	EX897699_12	Continuation (13 o

ALIGNMENTS

RESULT 1	AR428736	Sequence 16 from patent US 6642036.	684 bp	DNA	linear	PAT 18-DEC-2003
LOCUS	AR428736	Sequence 16 from patent US 6642036.	684 bp	DNA	linear	PAT 18-DEC-2003
DEFINITION	AR428736	Sequence 16 from patent US 6642036.	684 bp	DNA	linear	PAT 18-DEC-2003
ACCESSION	AR428736	Sequence 16 from patent US 6642036.	684 bp	DNA	linear	PAT 18-DEC-2003
VERSION	AR428736.1	GI:40188466				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 684)					
AUTHORS	Flint, D., Meyer, K. and Viitanen, P.					
TITLE	Sinapoylgucose:malate sinapoyltransferase form malate conjugates from benzoic acid glucosides					
JOURNAL	Patent: US 6642036-A 16 04-NOV-2003;					
FEATURES	WOX;	Location/Qualifiers				
source	1. .684	/organism="unknown"				
		/mol_type="genomic DNA"				
ORIGIN						
Alignment Scores:						
Pred. No.:		1.23e-83				684
Score:		887.00				170
Percent Similarity:		100.0%				Conservative: 0

```

Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 2 Gaps: 0

US-10-718-311-16 (1-170) x AR428736 (1-684)

Qy 1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20
Db 172 ATGCAGGTGTGGCATATGTCACACCCCGCGTTAACGCAACTCGCTGCGCTATTGT 231

Qy 21 LysGluIleProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40
Db 232 AAAGAGATCCCTGCCCTGGATCCGCAACTGCTCAGCTGGCTGTGCTGGAGGATCCATG 291

Qy 41 ThrLysArgPheGluGlnGlnGlyThrValSerValThrMetIleArgGluGlyPhe 60
Db 292 ACAAAACGTTTTTGAACAGCAGCGGAAAAACGGTAAGCGTGCAGCATGATCCGCAAGGGTTT 351

Qy 61 ValGluGlnAsnGluIleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrp 80
Db 352 GTCAGCAGAGATGAATAATCCCGAAGAACTGCGCGTGTCCGCAAGAGTCTCGTTACTGG 411

Qy 81 LeuArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100
Db 412 TTACGTGAATTTTGTATGTGCGGATGGTGAACCGTGGCTTGGCGGTCTGACCGTCGTT 471

Qy 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120
Db 472 CCTGTGTCAACGTTAAGCGGCGAGCTGGCGTTACAAAAATTGGGTAAACGCGTTA 531

Qy 121 GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAsp 140
Db 532 GGACGCTATCTGTTTCACATCATCGACATTACCCGGGACTTTATTGAGATAGGCGGTGAT 591

Qy 141 AlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 160
Db 592 GCCGGCTGTGGGGCGACGTTCCCGCTGGCATTAAGCGGTAAACCGCTGTGCTAACA 651

Qy 161 GluLeuPheLeuProAlaSerProLeuTyr 170
Db 652 GAACTGTTTTTACCGCGCTCACCGTTGTAC 681

RESULT 3
AX329368
LOCUS AX329368 684 bp DNA linear PAT 08-JAN-2002
DEFINITION Sequence 7 from Patent WO0194607.
ACCESSION AX329368
VERSION AX329368.1 GI:18102367
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1
AUTHORS Meyer, K., van Dyk, D.E. and Viitanen, P.V.
TITLE High level production of p-hydroxybenzoic acid in green plants
JOURNAL Patent: WO 0194607-A 7 13-DEC-2001;
FEATURES
Location/Qualifiers
source
1. .684
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="synthetic CPL"

ORIGIN
Alignment Scores:
Pred. No.: 1.23e-83 Length: 684
Score: 887.00 Matches: 170
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 2 Gaps: 0

US-10-718-311-16 (1-170) x AX329368 (1-684)

Qy 1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20
Db 172 ATGCAGGTGTGGCATATGTCACACCCCGCGTTAACGCAACTCGCTGCGCTATTGT 231

Qy 21 LysGluIleProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40
Db 232 AAAGAGATCCCTGCCCTGGATCCGCAACTGCTCAGCTGGCTGTGCTGGAGGATCCATG 291

Qy 41 ThrLysArgPheGluGlnGlnGlyThrValSerValThrMetIleArgGluGlyPhe 60
Db 292 ACAAAACGTTTTTGAACAGCAGCGGAAAAACGGTAAGCGTGCAGCATGATCCGCAAGGGTTT 351

Qy 61 ValGluGlnAsnGluIleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrp 80
Db 352 GTCAGCAGAGATGAATAATCCCGAAGAACTGCGCGTGTCCGCAAGAGTCTCGTTACTGG 411

Qy 81 LeuArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100
Db 412 TTACGTGAATTTTGTATGTGCGGATGGTGAACCGTGGCTTGGCGGTCTGACCGTCGTT 471

Qy 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120
Db 472 CCTGTGTCAACGTTAAGCGGCGAGCTGGCGTTACAAAAATTGGGTAAACGCGTTA 531

Qy 121 GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAsp 140
Db 532 GGACGCTATCTGTTTCACATCATCGACATTACCCGGGACTTTATTGAGATAGGCGGTGAT 591

Qy 141 AlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 160
Db 592 GCCGGCTGTGGGGCGACGTTCCCGCTGGCATTAAGCGGTAAACCGCTGTGCTAACA 651

Qy 161 GluLeuPheLeuProAlaSerProLeuTyr 170
Db 652 GAACTGTTTTTACCGCGCTCACCGTTGTAC 681

RESULT 2
AR455412
LOCUS AR455412 684 bp DNA linear PAT 20-FEB-2004
DEFINITION Sequence 7 from patent US 6683231.
ACCESSION AR455412
VERSION AR455412.1 GI:42689960
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 684)
AUTHORS Meyer, K., Viitanen, P.V. and Van Dyk, D.E.
TITLE High level production of p-hydroxybenzoic acid in green plants
JOURNAL Patent: US 6683231-A 7 27-JAN-2004;
FEATURES
Location/Qualifiers
source
1. .684
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores:
Pred. No.: 1.23e-83 Length: 684
Score: 887.00 Matches: 170
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 2 Gaps: 0

US-10-718-311-16 (1-170) x AR455412 (1-684)

```

232	AAAGAGATCCCTGCCTCGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCCATG	291
Qy	ThrLysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPhe	60
Db	ACAAAAAGTTTTGAACACACGCGAAAAACGGTAAGCGTGACGATGATCCGCGAAGGGTTT	351
Qy	ValGluGlnAanGluIleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrp	80
Db	GTCCAGCAGAAATGAATCCCCGAGAACTGCCGTGCTGCCGAAAGAGTCTCGTTACTGG	411
Qy	LeuArgGluIleLeuLeuCysAlaAspGlyGluProTrrLeuAlaGlyArgThrValVal	100
Db	TTACGTGAAATTTTGTATGTGCCGATGGTGAACCGTGGCTTCCGGTCTGACCGTCGTT	471
Qy	ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu	120
Db	CCTGTGTCAACGTTAAGCGGGCCGAGCTGGCGTTACAAAAATTCGGTAAAAACGCCGTTA	531
Qy	GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluLeuGlyArgAsp	140
Db	GGAGCGCTATCTGTTACATCATCATCGACATTAACCCGGGACTTTTATTGAGATAGGCCGTGAT	591
Qy	AlaGlyLeuTrpGlyArgGSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr	160
Db	GCCGGCTGTGGGGCGACGTTCCCGCCTGCGATTAAAGCGGTAAACCCGCTGTGCTAACA	651
Qy	GlutLeuPheLeuProAlaSerProLeuTyr	170
Db	GAACCTGTTTTTACCGGGGTCACCGTTGTAC	681

RESULT 5

A47932	A47932	495 bp	DNA	linear	PAT 07-MAR-1999
LOCUS	Sequence 1 from Patent WO9600788.				
DEFINITION	A47932				
ACCESSION	A47932.1	GI:2301795			
VERSION					
KEYWORDS	Escherichia coli				
SOURCE	Escherichia coli				
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.				
REFERENCE	1 (bases 1 to 495)				

```

AUTHORS      Heide, L., Severin, K. and Siebert, M.
TITLE        TRANSGENIC PLANTS WITH AN INCREASED
              PATENT: WO 9600788-A 11-JAN-1996;
              HEIDE-2092 (DE)
COMMENT       Other publication AU 2879295 960125
              Other publication DE 4423022 950524.
              Location/Qualifiers
FEATURES     source          1. .495
              /organism="Escherichia coli"
              /mol_type="unassigned DNA"
              /strain="MC4100"
              /db_xref="taxon:562"
ORIGIN

```

Alignment Scores:	
Pred. No.:	2.59e-80
Score:	854.00
Percent Similarity:	100.0%
Best Local Similarity:	100.0%
Query Match:	96.3%
DB:	2
Length:	495
Matches:	165
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-10-/18-311-16 (1-170) X A4/352 (1-495)

Qy	6	MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla	25
Db	1	ATGTCACACCCGGGTAAACGCACTGCGTGGCTATTGTTAAGAGATCCCTGCC	60
Qy	26	LeuAspProGlnLeuLeuAspTrrLeuLeuLeuGluAspSerMetThrLysArgPheGlu	45
Db	61	CTGGATCCGCAACTGCCTCGACTGCTGTTCGTGAGGATTCATGACAAAACGTTTGA	120



```
Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
Db 121 CAGCAGGGAACCGTAAGCGTACGATGATCGCGAAGGTTTGTTCGACGAGATGAA 180

Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
Db 181 ATCCCGGAAGAACTGCCGCTGCTCCGAAAGAGTCTCGTTACTCGTTACGTGAAATTTG 240

Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 105
Db 241 TTATGTCCGATGGTGAACCGTGGCTTCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTG 240

Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 105
Db 241 TTATGTCCGATGGTGAACCGTGGCTTCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTG 240

Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
Db 301 AGCGGGCGGAGCTGGCGTTACAAAAATTGGTAAACCGCGTTAGGACGCTATCTGTTC 360

Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
Db 361 ACATCATCGACATTAAACCCGGGACTTTTATGAGATAGCCGCTGATGCCGGCTGTGGGG 420

Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
Db 421 CGACGTTCCCGCTGCGATTAAAGCGTAAACCGCTGTTGCTAACAGAACTGTTTTACCG 480

Qy 166 AlaSerProLeuTyr 170
Db 481 GCGTCACCGTTGTAC 495

RESULT 6
AR428733 LOCUS 495 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 12 from patent US 6642036.
ACCESSION AR428733
VERSION AR428733.1 GI:40188463
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 495)
AUTHORS Flint, D., Meyer, K. and Viitanen, P.
TITLE Sinapoylglucose:malate sinapoyltransferase form malate conjugates
JOURNAL from benzoic acid glucosides
E. I. Du Pont de Nemours and Company; Wilmington;
WOX;

FEATURES
source Location/Qualifiers
1..495
/mol_type="genomic DNA"

ORIGIN
Alignment Scores:
Pred. No.: 2.59e-80 Length: 495
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.3% Indels: 0
DB: 2 Gaps: 0

US-10-718-311-16 (1-170) x AR428733 (1-495)

Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
Db 1 ATGTCACACCCCGCTTAAACGCACTGCGTGGCTGCGCTATTGTTAAAGAGATCCCTGCC 60

Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db 61 CTGGATCCGCAACTGCTCGACTGCGTGGCTGCGCTATTGTTGAGAGATCCAGCAAAACGTTTGA 120

Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
Db 121 CAGCAGGGAACCGTAAGCGTACGATGATCCGGAAGGTTTGTTCGACGAGATGAA 180

Qy 86 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
Db 181 ATCCCGGAAGAACTGCCGCTGCTCCGAAAGAGTCTCGTTACTCGTTACGTGAAATTTG 240

Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 105
Db 241 TTATGTCCGATGGTGAACCGTGGCTTCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTG 240
```

```
Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
Db 181 ATCCCGGAAGAACTGCCGCTGCTCCGAAAGAGTCTCGTTACTCGTTACGTGAAATTTG 240

Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 105
Db 241 TTATGTCCGATGGTGAACCGTGGCTTCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTG 240

Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
Db 301 AGCGGGCGGAGCTGGCGTTACAAAAATTGGTAAACCGCGTTAGGACGCTATCTGTTC 360

Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
Db 361 ACATCATCGACATTAAACCCGGGACTTTTATGAGATAGCCGCTGATGCCGGCTGTGGGG 420

Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
Db 421 CGACGTTCCCGCTGCGATTAAAGCGTAAACCGCTGTTGCTAACAGAACTGTTTTACCG 480

Qy 166 AlaSerProLeuTyr 170
Db 481 GCGTCACCGTTGTAC 495

RESULT 7
AR455409 LOCUS 495 bp DNA linear PAT 20-FEB-2004
DEFINITION Sequence 3 from patent US 6683231.
ACCESSION AR455409
VERSION AR455409.1 GI:42689957
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 495)
AUTHORS Meyer, K., Viitanen, P. V. and Van Dyk, D. E.
TITLE High level production of p-hydroxybenzoic acid in green plants
JOURNAL Patent: US 6683231-A 3 27-JAN-2004;
E. I. du Pont de Nemours and Company; Wilmington, DE

FEATURES
source Location/Qualifiers
1..495
/mol_type="genomic DNA"

ORIGIN
Alignment Scores:
Pred. No.: 2.59e-80 Length: 495
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.3% Indels: 0
DB: 2 Gaps: 0

US-10-718-311-16 (1-170) x AR455409 (1-495)

Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
Db 1 ATGTCACACCCCGCTTAAACGCACTGCGTGGCTGCGCTATTGTTAAAGAGATCCCTGCC 60

Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db 61 CTGGATCCGCAACTGCTCGACTGCGTGGCTGCGCTATTGTTGAGAGATCCAGCAAAACGTTTGA 120

Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
Db 121 CAGCAGGGAACCGTAAGCGTACGATGATCCGGAAGGTTTGTTCGACGAGATGAA 180

Qy 86 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
Db 181 ATCCCGGAAGAACTGCCGCTGCTCCGAAAGAGTCTCGTTACTCGTTACGTGAAATTTG 240

Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 105
Db 241 TTATGTCCGATGGTGAACCGTGGCTTCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTG 240
```



```
Db      241 TTATGTGCGGATGGTGAACCGTGGCTTGGCGGTGCTACCGTTCCTCTGTGCAACGTTA 300
Qy      106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
Db      301 AGCGGGCCGGAGTGGCTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 360
Qy      126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
Db      361 ACATCATCGACATTAAACCGGACTTTATTGAGATAGGCGGTATCCCGGGCTGTGGGG 420
Qy      146 ArgArgSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
Db      361 ACATCATCGACATTAAACCGGACTTTATTGAGATAGGCGGTATCCCGGGCTGTGGGG 420
Qy      146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
Db      421 CGACGTTCCCGCTCGGATTAGCGGTAAACCGCTGTTCGTAAACGAACTGTTTTTACCG 480
Qy      166 AlaSerProLeuTyr 170
Db      481 GCGTCACCGTTGTAC 495

RESULT 8
AX329364
LOCUS      AX329364
DEFINITION Sequence 3 from Patent WO0194607.
ACCESSION AX329364
VERSION    AX329364.1 GI:18102364
KEYWORDS   .
SOURCE     unidentified
           unclassified sequences.
ORGANISM   1
REFERENCE  1
AUTHORS    Meyer, K., van Dyk, D.E. and Viitanen, P.V.
TITLE      High level production of p-hydroxybenzoic acid in green plants
JOURNAL    Patent: WO 0194607-A 3 13-DEC-2001;
           E.I. Dupont De Nemours (US)
FEATURES   Location/Qualifiers
           source
             1..495
             /organism="unidentified"
             /mol_type="unassigned DNA"
             /db_xref="taxon:32644"
             /note="Description of Unknown Organism:E. coli"

ORIGIN
Alignment Scores:
Pred. No.:      2,598-80      Length:      495
Score:          854.00      Matches:    165
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    96.3%      Indels:      0
DB:             2          Gaps:        0

US-10-718-311-16 (1-170) x AX329364 (1-495)

Qy      6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
Db      1 ATGTACACCCCGCTTAACGCAACTGGCTGGCTATTTGTAAGAGATCCCTGCC 60
Qy      26 LeuAppProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db      61 CTGGATCCGCAACTGCTCGACTGGCTGTTCGTGGAGGATTCATGACAAAACGTTTGA 120
Qy      46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
Db      121 CAGCAGGGGAAAACCGTAAAGCGTGACGATGATCCCGAAGGGTTTGTGAGCAGAAATGAA 180
Qy      66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
Db      181 ATCCCCGGAAGAACTCCCGCTGCTGCCGAAAGAGTCTCGTTACTGTTAGTGAAATTTTG 240
Qy      86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 105
Db      241 TTATGTGCGGATGGTGAACCGTGGCTTGGCGGTGCTACCGTCTCTGTTGTCACAGTTA 300
Qy      106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
```

```
Db      301 AGCGGGCCGGAGTGGCTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 360
Qy      126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
Db      361 ACATCATCGACATTAAACCGGACTTTATTGAGATAGGCGGTATCCCGGGCTGTGGGG 420
Qy      146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
Db      421 CGACGTTCCCGCTCGGATTAGCGGTAAACCGCTGTTCGTAAACGAACTGTTTTTACCG 480
Qy      166 AlaSerProLeuTyr 170
Db      481 GCGTCACCGTTGTAC 495

RESULT 9
AX382254
LOCUS      AX382254
DEFINITION Sequence 12 from Patent WO0204653.
ACCESSION AX382254
VERSION    AX382254.1 GI:19577034
KEYWORDS   .
SOURCE     Escherichia coli
           Escherichia coli
           Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
           Enterobacteriaceae; Escherichia.
ORGANISM   1
REFERENCE  1
AUTHORS    Flint, D., Meyer, K. and Viitanen, P.V.
TITLE      Sinapolyglucose:maltate sinapolytransferase form malate conjugates
           from benzoic acid glucosides
JOURNAL    Patent: WO 0204653-A 12 17-JAN-2002;
           E.I. DUPONT DE NEMOURS AND COMPANY (US)
FEATURES   Location/Qualifiers
           source
             1..495
             /organism="Escherichia coli"
             /mol_type="unassigned DNA"
             /db_xref="taxon:562"

ORIGIN
Alignment Scores:
Pred. No.:      2,598-80      Length:      495
Score:          854.00      Matches:    165
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:    96.3%      Indels:      0
DB:             2          Gaps:        0

US-10-718-311-16 (1-170) x AX382254 (1-495)

Qy      6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
Db      1 ATGTACACCCCGCTTAACGCAACTGGCTGGCTATTTGTAAGAGATCCCTGCC 60
Qy      26 LeuAppProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db      61 CTGGATCCGCAACTGCTCGACTGGCTGTTCGTGGAGGATTCATGACAAAACGTTTGA 120
Qy      46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
Db      121 CAGCAGGGGAAAACCGTAAAGCGTGACGATGATCCCGAAGGGTTTGTGAGCAGAAATGAA 180
Qy      66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
Db      181 ATCCCCGGAAGAACTCCCGCTGCTGCCGAAAGAGTCTCGTTACTGTTAGTGAAATTTTG 240
Qy      86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 105
Db      241 TTATGTGCGGATGGTGAACCGTGGCTTGGCGGTGCTACCGTCTCTGTTGTCACAGTTA 300
Qy      106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
Db      301 AGCGGGCCGGAGTGGCTTACAAAATTGGGTAAACCGCTTAGGACGCTATCTGTTTC 360
```

Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145  
 Db 361 ACATCATCGACATTAAACCGGACTTTATTGAGATAGCGCGTGTGGGGG 420

Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165  
 Db 421 CGAGCTTCCCGCTGGGATTAGCGGTAAACCGCTGTGCTAACAGAACTGTTTTTACCG 480

Qy 166 AlaSerProLeuTyr 170  
 Db 481 CGGTACCGTTGTAC 495

RESULT 10  
 AR144788  
 LOCUS AR144788 854 bp DNA linear PAT 08-AUG-2001  
 DEFINITION Sequence 24 from patent US 6210937.  
 ACCESSION AR144788  
 VERSION AR144788.1 GI:15106655  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 UNCLASSIFIED.  
 REFERENCE 1 (bases 1 to 854)  
 AUTHORS Ward T.E., Watkins, C.S., Bulmer, D.K., Johnson, B.F. and Amarutunga, M.  
 TITLE Development of genetically engineered bacteria for production of selected aromatic compounds  
 JOURNAL Patent: US 6210937-A 24 03-APR-2001;  
 FEATURES Location/Qualifiers  
 source 1..854  
 /organism="unknown"  
 /mol\_type="unassigned DNA"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 4,67e-80 Length: 854  
 Score: 854.00 Matches: 165  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 96.3% Indels: 0  
 DB: 2 Gaps: 0

US-10-718-311-16 (1-170) x AR144788 (1-854)

Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25  
 Db 321 ATGTCACACCCGGGTTAAACGCACTGGCTGGCTGCCTATTGTAAAGAGATCCCTGCC 380

Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 45  
 Db 381 CTGGATCCGCAACTGCTCGACTGGCTGTGTGGAGGATTCATGACAAAAACGTTTGA 440

Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65  
 Db 441 CAGCAGGGAAGAAACGGTAGGCTGACGATGATCCGGAAGGTTGTTCGACAGAAATGA 500

Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85  
 Db 501 ATCCCGGAAGAACTGCCGCTGTCGCGAAGAGTCTCGTTACTGTTACGTTGAAATTTG 560

Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 105  
 Db 561 TTATGTGCGCATGGTGAACCGGTGGCTTCCCGGTCTGACCGTCTGTTGTGTCAACGTTA 620

Qy 106 SerGlyProGluLeuLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125  
 Db 621 ACGGGCCGAGCTGGCGTTACAAAATTTGGTAAACCGCTTAGGACGCTATCTGTT 680

Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145  
 Db 681 ACATCATCGACATTAAACCGGACTTTATTGAGATAGCGCGTGTGGGGG 740

Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165

Db 741 CGAGCTTCCCGCTGGGATTAGCGGTAAACCGCTGTGCTAACAGAACTGTTTTTACCG 800

Qy 166 AlaSerProLeuTyr 170  
 Db 801 CGGTACCGTTGTAC 815

RESULT 11  
 ECOUBIA  
 LOCUS 1498 bp DNA linear BCT 28-MAR-1994  
 DEFINITION Escherichia coli 4-hydroxybenzoate octaprenyl transferase (ubiA) gene, 5' end and chorismate lyase (ubiC) gene, complete cds.  
 ACCESSION M96268  
 VERSION M96268.1 GI:347886  
 KEYWORDS 4-hydroxybenzoate octaprenyl transferase; aerobic respiratory deficiency; chorismate lyase; membrane-bound protein; prenyltransferase; transferase; ubiA gene; ubiC gene; ubiquinone biosynthesis.  
 SOURCE Escherichia coli  
 ORGANISM Escherichia coli  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.  
 REFERENCE 1 (bases 1 to 1498)  
 AUTHORS Wu, G., Williams, H.D., Gibson, F. and Poole, R.K.  
 TITLE Mutants of Escherichia coli affected in respiration: the cloning and nucleotide sequence of ubiA, encoding the membrane-bound p-hydroxybenzoate:octaprenyltransferase  
 JOURNAL J. Gen. Microbiol. 139 (Pt 8), 1795-1805 (1993)  
 FEATURES PUBMED 8409922  
 Location/Qualifiers  
 source 1..1498  
 /organism="Escherichia coli"  
 /mol\_type="unassigned DNA"  
 /strain="K-12"  
 /db\_xref="taxon:562"  
 /map="91.8 min"  
 /tissue\_lib="cosmid"  
 <1..524  
 /note="ORF; the sequence before nucleotide number 110 might not be genuine. Three Clai sites, which is the enzyme used for cloning, were identified in this region, so it could be a cloning artifact"  
 /codon\_start=3  
 /transl\_table=11  
 /protein\_id="AAAI7026.1"  
 /db\_xref="GI:347887"  
 /translation="RFSEVNFDTQQMIDFINFNTKNKWLIDKNNNFQLIDQALRSTDMTKANVHLKIEWIRSDVSPIFETEDNLRFTNELTRNDNIFLPSVDDGPVMTVSSQRLHDLNAPTQDTNNSTVIYKSRHEMLPVNLTQETLFSSKSHGKVALPIFTASRAHRIMKGV"  
 RBS 735..738  
 gene 747..1244  
 /gene="ubiC"  
 CDS 747..1244  
 /gene="ubiC"  
 /codon\_start=1  
 /transl\_table=11  
 /product="chorismate lyase"  
 /protein\_id="AAAI7027.1"  
 /db\_xref="GI:347888"  
 /translation="MSPALPTQLRALRYCKEIPALDPQLLDMLLEDSTMKRFSSQOGKTVSVTMIREGFQNEIPSELPLLPKESRYLWREILLCADGEPWLAGRTVVPVSTLSGPELALQKLGKTLPLGRLFTSTLTRDFIEGRDAGLWGRRLRLSGKPLLTLELFLPASPLY"  
 gene 1257..1498  
 /gene="ubiA"  
 CDS 1257..1498  
 /gene="ubiA"  
 /note="The last part of this ORF is to be found in entry K00127.; ORF"  
 /codon\_start=1  
 /transl\_table=11  
 /product="4-hydroxybenzoate-octaprenyl transferase"

ECOURICA 2000 bp DNA linear BCT 26-APR-1993  
Escherichia coli chorismate lyase (ubiC), 4-hydroxybenzoate  
octaprenyl transferase (ubiA) genes, complete cds, and  
sn-glycerol-3-phosphate acyltransferase (plsB) genes, 3' end.  
M93136

```

VERSION M91316.1 GI:148099
KEYWORDS 4-hydroxybenzoate octaprenyl transferase; chorismate lyase; plbB
          gene; sn-glycerol-3-phosphate acyltransferase; ubiA gene; ubiC
          gene.
SOURCE   Escherichia coli W3110
ORGANISM Escherichia coli W3110
          Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
          Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 2000)
AUTHORS Lightner,V.A., Bell,R.M. and Modrich,P.
TITLE The DNA sequences encoding plbB and dgk loci of Escherichia coli
JOURNAL J. Biol. Chem. 258 (18), 10856-10861 (1983)
PUBMED 6309817
REFERENCE 2 (bases 1 to 2000)
AUTHORS Nichols,B.P. and Green,J.M.
TITLE Cloning and Sequencing of Escherichia coli ubiC and purification of
          chorismate lyase
JOURNAL Unpublished (1992)
COMMENT Original source text: Escherichia coli (sub_strain W3110, strain
          K-12) (library: Kohare lambda-lf8) DNA.
          Location/Qualifiers
            1..2000
              /organism="Escherichia coli W3110"
              /mol_type="genomic DNA"
              /strain="K-12"
              /sub_strain="W3110"
              /db_xref="taxon:316407"
              /map="91.8 min"
              /tissue_lib="Kohare lambda-lf8"
              380..877
                /gene="ubiC"
                380..877
                  /gene="ubiC"
                  /function="enzymatic chorismate --> p-hydroxybenzoate +
                    pyruvate"
                    /experiment="experimental evidence, no additional details
                      recorded"
                      /citation={2}
                      /codon_start=1
                      /transl_table=11
                      /product="chorismate lyase"
                      /protein_id="AAA24711.1"
                      /db_xref="GI:148100"
                      /translation="MSHPALTLQLRALRYCKEIPALDPOLLWLLEDSMTYKRFQOQK
                        TVSYMTIREGVEQNEIPEELPLPKESRYWLREILLCADGEPWLAKRTVPVSTLSG
                        PELALQKLGKTPGLRYLFTSLTRDFIEIGRDAGLMGRSRSLRSGKPLLLTELFLP
                        ASPLY"
                        890..1762
                          /gene="ubiA"
                          890..1762
                            /gene="ubiA"
                            /function="enzymatic - 3-octaprenyl-4-hydroxybenzoate
                              synthesis"
                              /experiment="experimental evidence, no additional details
                                recorded"
                                /citation={2}
                                /codon_start=1
                                /transl_table=11
                                /product="4-hydroxybenzoate-octaprenyl transferase"
                                /protein_id="AAA24712.1"
                                /db_xref="GI:148101"
                                /translation="MWSLSTQKLLAFHRLMRTDKPGALLLLWPTLWALWVATPGVP
                                  QLWTLAVPAGVWLMRAAGCVNDYADRFQHKVKTANRPLPSGAVTEKEARALFV
                                  LVLSFLVLTNTMTLLLSAALAAWVPFMKRYTHLPQVVLGAAPGWSIPMAFAA
                                  VSESPVLSCLMFLANILWAVDTQVAMVDKDDDKIGIKSTAILFGQYDKLIIGIL
                                  QIGVLAMAIIGELNGLWGYWSILVAGALFVYQQLIANREREACFKAPMNNYVG
                                  LVLEGLAMSYWHF"
                                  complement(1917..2000)
                                    /gene="plbB"
                                    complement(1917..2000)
                                      /gene="plbB"
                                      /EC_number="2.3.1.15"
                                      /function="lysophosphatidic acid synthesis"

```

---

```

/experiment="experimental evidence, no additional details
recorded"
/citation={1}
/codon_start=1
/transl_table=11
/product="sn-glycerol-3-phosphate acyltransferase"
/protein_id="AAA24713.1"
/db_xref="GI:148102"
/translation="MKVYQLLAELITSDVRLTIESATQGEG"

ORIGIN
Alignment Scores:
Pred. No.: 1,17e-79 Length: 2000
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.3% Indels: 0
DB: 15 Gaps: 0

US-10-718-311-16 (1-170) x ECOUBICA (1-2000)
Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProLa 25
   |||||
Db 380 ATGTCACACCCCGCGTTAAACGCAACTCGTGGCTGGGCTATTGTAAAGAGATCCCTGCC 439
Qy 26 LeuAspProGlnLeuLeuAspTyrPleuLeuLeuGluAspSerMetThrLysArgPheGlu 45
   |||||
Db 440 CTGGATCCGCAACTGCTCGACTGGCTGTGTGCTGGAGGATTCATGACAAAAACGTTTGA 499
Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
   |||||
Db 500 CAGCAGGGAANAACGGTAAGCGTGACGATGATCCGCGAAGGGTTTGTCCAGCAGAGATGA 559
Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeuLeu 85
   |||||
Db 560 ATCCCCGGAAGAACTGCCGCTGCTGCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTG 619
Qy 86 LeuCysAlaAspGlyGluProTyrPleuAlaGlyArgThrValValProValSerThrLeu 105
   |||||
Db 620 TTATGTCCGATGGTGAACCGTGGCTTGCCTGCTACCGCTGCTCTGTCTCAACGTTA 679
Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
   |||||
Db 680 AGCGGGCGGAGCTGGGGTTACAAAATTGGTAAACGCCGTTAGGACGCTATCTGTTC 739
Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluLeuGlyArgAspAlaGlyLeuTrpGly 145
   |||||
Db 740 ACATCATCGACATTAAACCCGGGACTTTATTGAGATAGCCCGTGATGCCGGCTGTGGGGG 799
Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
   |||||
Db 800 CGACGTTCCCGCCCTGCGATTAAAGCGGTAACCGCTGTTGTGTACAGAACTGTTTTACCG 859
Qy 166 AlaSerProLeuTyr 170
   |||||
Db 860 CGGTACCGTTGTATC 874

RESULT 14
ECOUBIPLS 2000 bp DNA linear BCT 26-APR-1993
LOCUS Escherichia coli 4-hydroxybenzoate octaprenyl transferase (ubiA)
DEFINITION gene complete cds, chorismate lyase (ubiC) gene complete cds,
          sn-glycerol-3-phosphate acyltransferase (plbB) gene, 3' end.
ACCESSION M93413.1 GI:148106
VERSION M93413
KEYWORDS 4-hydroxybenzoate octaprenyl transferase; chorismate lyase; plbB
          gene; sn-glycerol-3-phosphate acyltransferase; ubiA gene; ubiC
          gene.
SOURCE Escherichia coli W3110
ORGANISM Escherichia coli W3110
          Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
          Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 2000)

```

AUTHORS Lightner, V.A., Bell, R.M. and Modrich, P.  
 TITLE The DNA sequences encoding plxB and dgk loci of Escherichia coli  
 JOURNAL J. Biol. Chem. 258 (18), 10856-10861 (1983)  
 PUBMED 6309817  
 REFERENCE 2 (bases 1 to 2000)  
 AUTHORS Nichols, B.P. and Green, J.M.  
 TITLE Cloning and sequencing of Escherichia coli ubiC and purification of chorismate lyase  
 JOURNAL J. Bacteriol. 174 (16), 5309-5316 (1992)  
 PUBMED 1644758  
 COMMENT Original source text: Escherichia coli (sub\_strain W3110, strain X-12) (library: Kohara lambda) DNA.  
 FEATURES  
 source  
 1. .2000  
 /organism="Escherichia coli W3110"  
 /mol\_type="genomic DNA"  
 /strain="X-12"  
 /sub\_strain="W3110"  
 /db\_xref="taxon:316407"  
 /map="91.8 min"  
 /tissue\_lib="Kohara lambda"  
 380. .877  
 /gene="ubiC"  
 380. .877  
 /gene="ubiC"  
 /function="enzymatic chorismate --> p-hydroxybenzoate + pyruvate"  
 /experiment="experimental evidence, no additional details recorded"  
 /note="N-terminal amino acid sequence of chorismate lyase :NH2-SHPALQLRALRYCEIPAL-"  
 /citation={2}  
 /codon\_start=1  
 /transl\_table=11  
 /product="chorismate lyase"  
 /protein\_id="AAA24716.1"  
 /db\_xref="GI:148107"  
 /translation="MSHPALQLRALRYCKEIPALDPOLLDLLESDMTKRFEOQKG TVSTMIRGEFVEQNEIPEELPLPKESRYWLEILLCADGSPWLAGRTVVPVSLSG PELALQKLGKPLGRVLTSSLTLDFTIEIGRDAGLNGRSRLRSLSGPRLLTLELPL ASPLY"  
 890. .1762  
 /gene="ubiA"  
 890. .1762  
 /gene="ubiA"  
 /function="enzymatic - 3-octaprenyl-4-hydroxybenzoate synthesis"  
 /experiment="experimental evidence, no additional details recorded"  
 /citation={2}  
 /number=2  
 /codon\_start=1  
 /transl\_table=11  
 /product="4-hydroxybenzoate-octaprenyl transferase"  
 /protein\_id="AAA24717.1"  
 /db\_xref="GI:148108"  
 /translation="MEWSLTQNKLIAFHRLMETHDKPIGALLLLWPTLWLVATPGVP QLWILAVFAGVWLNRAAGCVNDYADRFQGHVVRTPANPLPSGAVTEKRALFVV LVLISFLVLTUNTTILLTAALAWYIPMKRYTHLPQVVLGAAGFWSIPMAFAA VSESPLSCWLMFLANLWAVDTQYAMVDVDDDKVIGIKSTAILFGQYDKLIIGIL QIGVLALMAIIGELNGLGWYYSILVAGALFVYQOKLIANRERACFKAFMNNYVG LVLFLGLAMSYWHF"  
 complement (1917. .2000)  
 /gene="plSB"  
 complement (1917. .2000)  
 /gene="plSB"  
 /EC\_number="2.3.1.15"  
 /function="lysophosphatidic acid synthesis"  
 /experiment="experimental evidence, no additional details recorded"  
 /citation={1}  
 /codon\_start=1  
 /transl\_table=11

/product="sn-glycerol-3-phosphate acyltransferase"  
 /protein\_id="AAA24718.1"  
 /db\_xref="GI:148109"  
 /translation="MKVYQLLAELITSVRLTIESATQGEG"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,17e-79 Length: 2000  
 Score: 854.00 Matches: 165  
 Percent Similarity: 100.0% Conservatives: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Indels: 0  
 Query Match: 96.3%  
 DB: 15 Gaps: 0  
 US-10-718-311-16 (11-170) x ECOUBIPLS (1-2000)  
 QY 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 25  
 Db 380 ATGTACACCCCGCGTTAACGCAACTGCGTGCCTGCGCTATTGTAAGAGATCCCTGCC 439  
 QY 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45  
 Db 440 CTGGATCCGCAACTGCTGACTGGCTGTGCTGGAGGATCCATGACAAAACGTTTGA 499  
 QY 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65  
 Db 500 CAGCAGGGGAAACCGTAAGCGTGACGATGATCCGGAAGGGTTGTCGAGCAGATGAA 559  
 QY 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeuLeu 85  
 Db 560 ATCCCGAAGAACTCGCTGCTGCCGAAAGAGTCTCGTTACTGTTAGCTGAAATTTTG 619  
 QY 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 105  
 Db 620 TTATGTGCGGATGGTGAACCGGTGCTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTG 679  
 QY 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125  
 Db 680 AGCGGGCGGAGCTGCGCTTACAAAATTTGGTAAACCGCTTAGGACGCTATCTGTTTC 739  
 QY 126 ThrSerSerThrLeuThrArgAspPheIleGluLeuGlyArgAspAlaGlyLeuTrpGly 145  
 Db 740 ACATCATCGACATTAAACCGGACTTTATTGAGATAGGCGGTGATGCGCGGCTGTGGGG 799  
 QY 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165  
 Db 800 CGACGTTCCCGCTCGGATTAGCGTAAACCGCTGTGCTAACAGAACTGTTTTTACCG 859  
 QY 166 AlaSerProLeuTyr 170  
 Db 860 CGGTACCGTTGTAC 874  
 RESULT 15  
 ECUBIAC  
 LOCUS  
 DEFINITION E.coli ubiC and ubiA genes for chorismate lyase and 4-hydroxybenzoate octaprenyltransferase.  
 ACCESSION X57434  
 VERSION X57434.1 GI:43233  
 KEYWORDS 4-hydroxybenzoate octaprenyl transferase; 4-hydroxybenzoate synthetase; chorismate lyase.  
 SOURCE Escherichia coli  
 ORGANISM Escherichia coli  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.  
 REFERENCE 1  
 AUTHORS Nishimura, K., Nakahigashi, K. and Inokuchi, H.  
 TITLE Location of the ubiA gene on the physical map of Escherichia coli  
 JOURNAL J. Bacteriol. 174 (17), 5762 (1992)  
 PUBMED 1512213  
 REFERENCE 2 (bases 1 to 2034)  
 AUTHORS Nishimura, K.  
 TITLE Direct Submission

JOURNAL Submitted (28-JAN-1991) K. Nishimura, Faculty of Sciences, Kyoto  
University, Sakyo-ku, Kyoto 606, Japan

FEATURES  
source

1. .2034  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/strain="K12 W3110"  
/db\_xref="taxon:562"  
/map="91.7 min"  
/clone\_lib="Kohara library"  
137. .142  
159. .164  
291. .899  
/gene="ubiC"  
291. .899  
/gene="ubiC"  
/codon\_start=1  
/transl\_table=11  
/product="4-hydroxybenzoate synthetase"  
/protein\_id="CAA40681.1"  
/db\_xref="GI:43234"  
/db\_xref="GOA:P26602"  
/translation="MRLRFCCVLDHLICFTSPVNTFLRYNAFTLCNGFCQMSHPALT  
QLRALRYCKEIPALDPQLDMLLEDSTKBEQOGKTVSVTMIREFGVEQNEIPEEL  
PULPKESYWIREEILLCADGEFWLAGRVVPVSTLSGPELALQKLKLTPLGRYLTSS  
TLTRDFEIGRDAGLWGRRRSLRUSGKPLLLTEUFLPASPLY"  
912. .1784  
/gene="ubiA"  
912. .1784  
/gene="ubiA"  
/codon\_start=1  
/transl\_table=11  
/product="4-hydroxybenzoate octaprenyltransferase"  
/protein\_id="CAA40682.1"  
/db\_xref="GI:43235"  
/db\_xref="GOA:P26601"  
/translation="MWSLTKQKLLAFHLMRTDKPIGALLLLWPTLWALMVATPGVP  
QLWILAVFVAGVLMRAAGCVVNDYADRFQGHVKRTANRPLPSGAVTEKEARALFVV  
LVLSIFLVLTNTMTILLSAALAWVPFMKRYTHLPQVILGAAFGWSIPWAFAA  
VSESPLSCWLMFLANIWAVYDQYAMVDRDDVKIGIKSTAILFGYDKLIIGIL  
QIGVLALMAIGELNLGWGYWSILVAGALFVYQQLIANRERECFKAPMNNYVG  
LVFLGLANSIWHF"

-35\_signal  
-10\_signal  
gene  
CDS

gene  
CDS

ORIGIN

Alignment Scores:  
Pred. No.: 1.2e-79 Length: 2034  
Score: 854.00 Matches: 165  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 96.3% Indels: 0  
DB: 15 Gaps: 0

US-10-718-311-16 (1-170) x ECUBIAC (1-2034)

Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25  
Db 402 ATGTCACACCCCGCGTTAAACGCACTGCGTGGCTATTGTAAAGAGATCCCTGCC 461  
Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45  
Db 462 CTGGATCCGCAACTGCTCGACTGGCTGTGTGGAGGATTCATGACAAAAACGTTTGA 521  
Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAsnGlu 65  
Db 522 CAGCAGGGGAAAAACGGTAAGCGTGNCGATGATCCCGAAGGGTTTGTCCAGCAGAAATGA 581  
Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85  
Db 582 ATCCCGGAAGAACTGCCGCTGCTGCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTG 641

Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 105  
Db 642 TTAITGTCGGATGGTGAAACCGTGGCTTCCCGGTGTTACCGTCGTTCTGTGTCAACGTTA 701  
Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrIleuPhe 125  
Db 702 AGCGGGCGGAGCTGGCGTTACAAAAAATTGGGTAAACGCCGTTAGGACGCTATCTGTTT 761  
Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145  
Db 762 ACATCATCGACATTAAACCCGGGACTTTATTGAGATAGCCGCGTGTGCGGCTGTGGGG 821  
Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165  
Db 822 CGACGTTCCCGCTGCGATTAAAGCGGTAAACCGCTGTGTCTTAACAGAACTGTTTTTACCG 881  
Qy 166 AlaSerProLeuTyr 170  
Db 882 GCGTCACCGTTGTAC 896

Search completed: May 30, 2006, 05:52:57  
Job time : 5159.27 secs

GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 29, 2006, 11:01:13 ; Search time 207.045 Seconds  
(without alignments)  
2304.491 Million cell updates/sec

Title: US-10-718-311-16

Perfect score: 887

Sequence: 1 MQVWMSHPALQRLALRYC.....RLSGKPLLTLLFLPASPLY 170

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1403666 seqs, 935554401 residues

Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-O=/abs8/ABSSWEB spool/US10718311/runat 26052006.164922.13534/app\_query.fasta\_1  
-DB=issued\_Patents\_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-HOST=abs802p -USER=US10718311 @CGN 1.1 615 @runat 26052006.164922.13534  
-NCPU=6 -ICPU=3 -NO MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/ina/1 COMB.seq.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/ina/5 COMB.seq.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/ina/6A COMB.seq.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/ina/6B COMB.seq.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/ina/7 COMB.seq.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/ina/H COMB.seq.\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/2/ina/PCUTUS COMB.seq.\*
- 8: /EMC\_Celerra\_SIDS3/ptodata/2/ina/PP COMB.seq.\*
- 9: /EMC\_Celerra\_SIDS3/ptodata/2/ina/RE COMB.seq.\*
- 10: /EMC\_Celerra\_SIDS3/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	887	100.0	684	3	US-09-896-866B-16
2	887	100.0	684	3	Sequence 16, Appl
3	854	96.3	495	3	Sequence 7, Appl
4	854	96.3	495	3	Sequence 12, Appl
5	854	96.3	495	3	Sequence 3, Appl
6	854	96.3	6541	3	Sequence 24, Appl
7	673	75.9	645	3	Sequence 25, Appl
8	401	45.2	537	3	Sequence 1234, Ap

9	186	21.0	630	3	US-09-252-991A-7442	Sequence 7442, Ap	
10	174.5	19.7	519	3	US-09-540-236-1889	Sequence 1889, Ap	
11	171.5	19.3	269223	3	US-09-596-002-41	Sequence 41, Appl	
12	159	17.9	540	3	US-09-328-352-343	Sequence 343, Appl	
13	146.5	16.5	1248	3	US-09-252-991A-7361	Sequence 7361, Ap	
c	14	121	13.6	195	3	US-09-543-681A-3086	Sequence 3086, Ap
15	90	10.1	2127	3	US-09-818-780-75	Sequence 75, Appl	
16	85	9.6	888	3	US-09-107-433-1721	Sequence 1721, Ap	
c	17	83	9.4	84839	3	US-09-949-016-15816	Sequence 15816, A
18	82	9.2	601	3	US-09-949-016-69693	Sequence 69693, A	
19	82	9.2	807	3	US-09-583-110-1232	Sequence 1232, Ap	
20	81	9.1	1176	3	US-09-252-991A-10640	Sequence 10640, A	
21	81	9.1	1581	3	US-09-252-991A-10731	Sequence 10731, A	
c	22	81	2055	3	US-09-252-991A-10433	Sequence 10433, A	
c	23	81	62354	3	US-09-949-016-16188	Sequence 16188, A	
c	24	80	6463	2	US-08-962-284-3	Sequence 3, Appl	
c	25	80	68778	3	US-09-949-016-16406	Sequence 16406, A	
c	26	80	72455	3	US-09-949-016-13793	Sequence 13793, A	
c	27	80	77535	3	US-09-949-016-14279	Sequence 14279, A	
c	28	80	77535	3	US-09-949-016-14280	Sequence 14280, A	
c	29	80	77535	3	US-09-949-016-14281	Sequence 14281, A	
c	30	80	10475	3	US-09-949-016-12115	Sequence 12115, A	
c	31	80	111282	3	US-09-754-250-3	Sequence 3, Appl	
c	32	80	111282	3	US-10-094-989-3	Sequence 3, Appl	
c	33	79.5	9.0	47981	3	US-09-679-279-1	Sequence 1, Appl
c	34	79.5	9.0	4403765	3	US-09-103-840A-2	Sequence 2, Appl
c	35	79.5	9.0	4411529	3	US-09-103-840A-1	Sequence 1, Appl
c	36	79	8.9	5246	3	US-09-799-451-474	Sequence 474, Appl
c	37	78.5	8.9	1486	2	US-08-250-859-1	Sequence 1, Appl
c	38	78.5	8.9	1486	2	US-08-347-029-1	Sequence 1, Appl
c	39	78.5	8.9	1486	2	US-08-490-803-1	Sequence 1, Appl
c	40	78.5	8.9	1486	2	US-08-457-254-1	Sequence 1, Appl
c	41	78.5	8.9	1486	2	US-08-484-257-1	Sequence 1, Appl
c	42	78.5	8.9	1486	3	US-08-999-927-3	Sequence 3, Appl
c	43	78.5	8.9	1486	3	US-08-461-819-3	Sequence 1, Appl
c	44	78.5	8.9	1486	7	PCT-US94-08806-1	Sequence 1, Appl
c	45	78.5	8.9	1486	7	PCT-US95-01775-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-09-896-866B-16  
; Sequence 16, Application US/09896866B  
; Patent No. 6642036  
; GENERAL INFORMATION:  
; APPLICANT: Flint, Dennis  
; APPLICANT: Meyer, Knut  
; APPLICANT: Viitanen, Paul  
; TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugate  
; FILE REFERENCE: BCI034 US NA  
; CURRENT APPLICATION NUMBER: US/09/896,866B  
; CURRENT FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: 60/216,615  
; PRIOR FILING DATE: 2000-07-07  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 16  
; LENGTH: 684  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; NAME/KEY: misc.feature  
; OTHER INFORMATION: open reading frame of the chloroplast-targeted CPL fusion protein  
US-09-896-866B-16

Alignment Scores:  
Pred. No.: 1,94e-100 Length: 684  
Score: 887.00 Matches: 170  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0



```
DB: 3 Gaps: 0
US-10-718-311-16 (1-170) x US-09-896-866B-16 (1-684)
Qy 1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20
Db 172 ATGCAGGTGTGGCATATGTACACCCCGGTTAACGCAACTGCGTGGCTATTTGT 231
Qy 21 LysGluileProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40
Db 172 ATGCAGGTGTGGCATATGTACACCCCGGTTAACGCAACTGCGTGGCTATTTGT 231
Qy 21 LysGluileProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40
Db 232 AAAGAGATCCCTGCGCTGCACTGCTCGACTGCGTGTGCTGGAGGATTCATG 291
Qy 41 ThrLysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPhe 60
Db 292 ACAAAACGTTTTTGAACACAGCAGGAGGAAAAACGGTAAGCGTGCATCCCGGAAAGGGTTT 351
Qy 61 ValGluGlnAanGluileProGluLeuProLeuLeuProLysGluSerArgTyrTrp 80
Db 352 GTCGAGCAGATGAATCCCGAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGG 411
Qy 81 LeuArgGluileLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100
Db 412 TTACGTGAATTTTGTATGTGCGCATGCTGAACCGTGGCTTGCCGGTCTGACCGTCGT 471
Qy 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120
Db 472 CCTGTGTCACCGTTAAGCGGGCGGAGCTGCGGTTACAAAAATTGGGTAAACCGCGTTA 531
Qy 121 GlyArgTyrLeuPheThrSerThrLeuThrArgAspPheIleGluileGlyArgAsp 140
Db 532 GGACGCTATCTGTTACATCATCGACATTAAACCGGGACTTTATTTGAGTAGGCGCGTAT 591
Qy 141 AlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 160
Db 592 GCCGGGCTGTGGGGCGAGCTTCCCGCTCGGATTAAACCGTTAAACCGCTGTGTGTAACA 651
Qy 161 GluLeuPheLeuProAlaSerProLeuTyr 170
Db 652 GAACGTGTTTTTACCGGGCTCACCCTGTGTAC 681
RESULT 2
US-09-855-341-7
; Sequence 7, Application US/09855341
; Patent No. 6683231
; GENERAL INFORMATION:
; APPLICANT: VIITANEN, PAUL V.
; APPLICANT: MEYER, KNUT
; APPLICANT: VAN DYK, DREW
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID
; TITLE OF INVENTION: IN GREEN PLANTS
; FILE REFERENCE: BC1015 US NA
; CURRENT APPLICATION NUMBER: US/09/855,341
; CURRENT FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 7
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:synthetic CPL
US-09-855-341-7
Alignment Scores:
Pred. No.: 1,948-100 Length: 684
Score: 887.00 Matches: 170
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 3 Gaps: 0
US-10-718-311-16 (1-170) x US-09-855-341-7 (1-684)
```

```
Qy 1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20
Db 172 ATGCAGGTGTGGCATATGTACACCCCGGTTAACGCAACTGCGTGGCTATTTGT 231
Qy 21 LysGluileProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40
Db 232 AAAGAGATCCCTGCGCTGGAATCCGCAACTGCTGCATGGCTGTTGCTGGAGGATTCATG 291
Qy 41 ThrLysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPhe 60
Db 292 AAAAAACGTTTTTGAACACAGCAGGAGGAAAAACGGTAAGCGTGCATCCCGGAAAGGGTTT 351
Qy 61 ValGluGlnAanGluileProGluLeuProLeuLeuProLysGluSerArgTyrTrp 80
Db 352 GTCGAGCAGATGAATCCCGAGAACTGCGCTGCTGCCGAAAGAGTCTCGTTACTGG 411
Qy 81 LeuArgGluileLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100
Db 412 TTACGTGAATTTTGTATGTGCGCATGCTGAACCGTGGCTTGCCGGTCTGACCGTCGT 471
Qy 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120
Db 472 CCTGTGTCACCGTTAAGCGGGCGGAGCTGCGGTTACAAAAATTGGGTAAACCGCGTTA 531
Qy 121 GlyArgTyrLeuPheThrSerThrLeuThrArgAspPheIleGluileGlyArgAsp 140
Db 532 GGACGCTATCTGTTACATCATCGACATTAAACCGGGACTTTATTTGAGTAGGCGCGTAT 591
Qy 141 AlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 160
Db 592 GCCGGGCTGTGGGGCGAGCTTCCCGCTCGGATTAAACCGTTAAACCGCTGTGTGTAACA 651
Qy 161 GluLeuPheLeuProAlaSerProLeuTyr 170
Db 652 GAACGTGTTTTTACCGGGCTCACCCTGTGTAC 681
RESULT 3
US-09-896-866B-12
; Sequence 12, Application US/09896866B
; Patent No. 6642036
; GENERAL INFORMATION:
; APPLICANT: Flint, Dennis
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugates
; TITLE OF INVENTION: Benzoic Acid Glucosides
; FILE REFERENCE: BC1034 US NA
; CURRENT APPLICATION NUMBER: US/09/896,866B
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/216,615
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 12
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-896-866B-12
Alignment Scores:
Pred. No.: 1,48e-96 Length: 495
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.3% Indels: 0
DB: 3 Gaps: 0
US-10-718-311-16 (1-170) x US-09-896-866B-12 (1-495)
Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysGluileProAla 25
Db 1 ATGTCACACCCCGGTTAAACGCAACTGCGTGGCTGCGCTATTGTTAAAGAGATCCCTGCC 60
```



Qy	26	LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrIysArgPheGlu	45
Db	61	CTGGATCCGCACCTCTCGACTGGCTGTTTGGTGGAGGATTCATGACAAACGTTTTTGAA	120
Qy	46	GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu	65
Db	121	CAGCAGGGGAAAAACGGTAAGCGTGACGATATCCGCGAAGGGTTGTCCAGCAGATGAA	180
Qy	66	IleProGluGluLeuProLeuLeuProIysGluSerArgTyrTrpLeuArgGluIleLeu	85
Db	181	ATCCCCGAAGAACTCCCGCTGCTGCCGAAGAAGTCTCGTTACTGGTTACGTGAAATTTTG	240
Qy	86	LeuCyAsAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu	105
Db	241	TTATGTGCCGATGGTGAACCGTGGCTTCCCGCTCGTACCGTCTCTGTGCAACGTTA	300
Qy	106	SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe	125
Db	301	AGCGGGCCGAGCTGGCGTTACAAAAATGGGTAAAAACGCCGTAGGACGCTATCTGTTTC	360
Qy	126	ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly	145
Db	361	ACATCATCGCATTAACCCGGGACTTTATTTAGATAGCCCGTGATGCCGGGCTGTGGGG	420
Qy	146	ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro	165
Db	421	CGAGTTCCCGCTCGGATTAACCGGTAAACCGCTGTGTCTAACAGACTGTTTTTACC	480
Qy	166	AlaSerProLeuTyr	170
Db	481	CGGTACACCGTTGTAC	495

RESULT 4  
US-09-855-341-3  
; Sequence 3, Application US/09855341

```

1 Patent No. 5685231
2
3 GENERAL INFORMATION:
4
5 APPLICANT: VIITANEN, PAUL V.
6
7 APPLICANT: MEYER, KNUIT
8
9 APPLICANT: VAN DYK, DREW
10
11 TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID
12
13 TITLE OF INVENTION: IN GREEN PLANTS
14
15 FILE REFERENCE: BC1015 US NA
16
17 CURRENT APPLICATION NUMBER: US/09/855,341
18
19 CURRENT FILING DATE: 2001-05-15
20
21 NUMBER OF SEQ ID NOS: 16
22
23 SOFTWARE: MICROSOFT OFFICE 97
24
25 SEQ ID NO 3
26
27 LENGTH: 495
28
29 TYPE: DNA
30
31 ORGANISM: Unknown Organism
32
33 FEATURE:
34
35 OTHER INFORMATION: Description of Unknown Organism:E. coli
36
37 US-09-855-141-3

```

Alignment Scores:	
Pred. No.:	1.48e-96
Score:	854.00
Percent Similarity:	100.0%
Best Local Similarity:	100.0%
Query Match:	96.3%
DB:	3
Gaps:	0
Indels:	0
Mismatches:	0
Conservative:	0
Matches:	165
Length:	495

US-10-718-311-16 (1-170) x US-09-855-341-3 (1-495)

Qy	6	MetSerHisProAlaLeuThrClnLeuArgAlaLeuArgTyrCysIysGluIleProAla	25
Db	1	ATGTCACACCCCGCGTTTAAACGCAATGGCGTGGCGCTATTGTAAAGAGATCCCTGCC	60
Qy	26	LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrIysArgPheGlu	45
Db	61	CTGGATCGCAACTCCTCGACTGGCTGTTCTGGAGGATTCCATGACAAACGTTTGA	120

Qy	46	GlnGlnGlyIysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu	65
Db	121	CAGCAGGGAAAAACGGTAAAGCGTGACGATGATCCGCAAGGGGTTTGTCGAGCAGAAATGAA	180
Qy	66	IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu	85
Db	181	ATCCCCGAAGAACTGCCGCTGTCGCCGAAGAGTCTCGTTACTGGTTACGTGAAATTTTG	240
Qy	86	LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu	105
Db	241	TTATGTGCCGATGGTGAACCGTGGCTTCCCGCTCGTACCGTCTCTGTGTCACAGTTTA	300
Qy	106	SerGlyProGluLeuAlaLeuGlnLysIeuGlyLysThrProLeuGlyArgTyrLeuPhe	125
Db	301	AGCGGGCCGAGCTGGCGTTTACAAAAATTGGGTAAAAACGCCCTTACGACGCTATCTGTTC	360
Qy	126	ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly	145
Db	361	ACATCATCGACATTAACCCGGACCTTATTGAGATAGCCCGTGATGCCGGGCTGTGGGG	420
Qy	146	ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro	165
Db	421	CGAGTTCGCCGCTGGGATTAAACGGGTAAACCGCTGTGTGCTAACAGAACTGTTTTTACC	480
Qy	166	AlaSerProLeuTyr	170
Db	481	CGGTACCGTTGTAC	495

5 TILISA

```

00001 3
00002  US-09-064-693A-24
00003  ; Sequence 24, Application US/09064693A
00004  ; Patent No. 6210937
00005  ; GENERAL INFORMATION:
00006  ; APPLICANT: Ward, Thomas E.
00007  ; TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY
00008  ; TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION
00009  ; TITLE OF INVENTION: OF A SPECIFIC PLASTICS PRECURSOR
00010  ; NUMBER OF SEQUENCES: 26
00011  ; CORRESPONDENCE ADDRESS:
00012  ; ADDRESSEE: W. Gary Goodson
00013  ; ADDRESSEE: INEL--Lockheed Martin Idaho
00014  ; ADDRESSEE: Technologies Co.
00015  ; STREET: P.O. Box 1625
00016  ; CITY: Idaho Falls
00017  ; STATE: Idaho
00018  ; COUNTRY: USA
00019  ; ZIP: 83415-3810
00020  ; COMPUTER READABLE FORM:
00021  ; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB storage
00022  ; COMPUTER: Toshiba Satellite Pro T2150CDS
00023  ; OPERATING SYSTEM: Windows95
00024  ; SOFTWARE: Word Perfect 7.0
00025  ; CURRENT APPLICATION DATA:
00026  ; APPLICATION NUMBER: US/09/064,693A
00027  ; FILING DATE:
00028  ; CLASSIFICATION: 435
00029  ; PRIOR APPLICATION DATA:
00030  ; APPLICATION NUMBER:
00031  ; FILING DATE:
00032  ; ATTORNEY/AGENT INFORMATION:
00033  ; NAME: W. Gary Goodson
00034  ; REGISTRATION NUMBER: 22,387
00035  ; REFERENCE/DOCKET NUMBER: LIT-PI-296
00036  ; TELECOMMUNICATION INFORMATION:
00037  ; TELEPHONE: (208)526-9469
00038  ; TELEFAX: (208)526-8339
00039  ; INFORMATION FOR SEQ ID NO: 24:
00040  ; SEQUENCE CHARACTERISTICS:
00041  ; LENGTH: 854 base pairs
00042  ; TYPE: nucleic acid
00043  ; STRANDEDNESS: double
00044  ; TOPOLOGY: linear
00045  US-09-064-693A-24

```

```
Alignment Scores:
Pred. No.: 3,396-96 Length: 854
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.3% Indels: 0
DB: 3 Gaps: 0

US-10-718-311-16 (1-170) x US-09-064-693A-24 (1-854)
Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
Db 321 ATGTCACACCCCGGTTAACGCACTGCGTGGCTGCGCTATTGTAAGAGATCCCTGCC 380
Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db 381 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGA 440
Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
Db 441 CAGCAGGGGAAAAACGGTAAGCGTGACGATGATCCGCGAAGGTTTGTCCGACAGAAATGA 500
Qy 66 IleProGluGluLeuLeuProLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
Db 501 ATCCCGGAAGAACTGCGCTGCTGCCGAAAGAGTCTGTTACTGGTTACGTGAATTTTG 560
Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrIleu 105
Db 561 TTATGTGCGGATGGTGAACCGTGGCTTCCGCTGCTACCGTCTGTTCTGTCAACGTTA 620
Qy 106 SerGlyProGluLeuAlaLeuGlnLysGluLysThrProLeuGlyArgTyrIleuphe 125
Db 621 AGCGGCGGAGCTGGCGTTACAAAATTTGGTAAACCGCGTTAGGACGCTATCTGTTC 680
Qy 126 ThrSerSerThrIleuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
Db 681 ACATCATCGACATTAACCCGGGACTTTATTAGATAGGCCGTGATGCCGGCTGTGGGG 740
Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
Db 741 CGAGTTCCCGCTGCGATTAAGCGTTAAACCGCTTTGCTTAACAGAACTGTTTTTACCG 800
Qy 166 AlaSerProLeuTyr 170
Db 801 CGCTCACCGTTGTAC 815

RESULT 6
US-09-064-693A-25
; Sequence 25, Application US/09064693A
; Patent No. 6210937
; GENERAL INFORMATION:
; APPLICANT: Ward, Thomas E.
; TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY
; TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION
; TITLE OF INVENTION: OF A SPECIFIC PLASTICS PRECURSOR
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W. Gary Goodson
; ADDRESSEE: INEEL--Lockheed Martin Idaho
; ADDRESSEE: Technologies Co.
; STREET: P.O. Box 1625
; CITY: Idaho Falls
; STATE: Idaho
; COUNTRY: USA
; ZIP: 83415-3810
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: Toshiba Satellite Pro T2150CDS
; OPERATING SYSTEM: Windows95
; SOFTWARE: Word Perfect 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/064,693A

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: W. Gary Goodson
; REGISTRATION NUMBER: 22,387
; REFERENCE/DOCKET NUMBER: LIT-PI-296
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (208)526-9469
; TELEFAX: (208)526-8339
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6641 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-064-693A-25

Alignment Scores:
Pred. No.: 7,76e-95 Length: 6641
Score: 854.00 Matches: 165
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.3% Indels: 0
DB: 3 Gaps: 0

US-10-718-311-16 (1-170) x US-09-064-693A-25 (1-6641)
Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
Db 321 ATGTCACACCCCGGTTAACGCACTGCGTGGCTGCGCTATTGTAAGAGATCCCTGCC 380
Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db 381 CTGGATCCGCAACTGCTCGACTGGCTGTTGCTGGAGGATTCATGACAAAACGTTTGA 440
Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
Db 441 CAGCAGGGGAAAAACGGTAAGCGTGACGATGATCCGCGAAGGTTTGTCCGACAGAAATGA 500
Qy 66 IleProGluGluLeuLeuProLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
Db 501 ATCCCGGAAGAACTGCGCTGCTGCCGAAAGAGTCTGTTACTGGTTACGTGAATTTTG 560
Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrIleu 105
Db 561 TTATGTGCGGATGGTGAACCGTGGCTTCCGCTGCTACCGTCTGTTCTGTCAACGTTA 620
Qy 106 SerGlyProGluLeuAlaLeuGlnLysGluLysThrProLeuGlyArgTyrIleuphe 125
Db 621 AGCGGCGGAGCTGGCGTTACAAAATTTGGTAAACCGCGTTAGGACGCTATCTGTTC 680
Qy 126 ThrSerSerThrIleuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly 145
Db 681 ACATCATCGACATTAACCCGGGACTTTATTAGATAGGCCGTGATGCCGGCTGTGGGG 740
Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
Db 741 CGAGTTCCCGCTGCGATTAAGCGTTAAACCGCTTTGCTTAACAGAACTGTTTTTACCG 800
Qy 166 AlaSerProLeuTyr 170
Db 801 CGCTCACCGTTGTAC 815

RESULT 7
US-09-489-039A-1234
; Sequence 1234, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
```



```
US-10-718-311-16 (1-170) x US-09-252-991A-7442 (1-630)
Qy 22 GluileProAlaLeuAaspProGlnLeuLeuAaspTrpLeuLeuGluAaspSerMetThr 41
Db 127 CAACGTGTCAGAGCGTCTCCCGCGCTCTCGACTGGCTGTCGAGGGGTGCTGAC 186
Qy 42 LysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluPheVal 61
Db 187 CGGCGC-----CTGACCGCCTGGCGCGCGCGGCTTC 219
Qy 62 GluGlnAenGluileProGluLeuLeuProLeuLeu----- 73
Db 220 CGCGTCGACCCCTGCTGAAGGTGGCGAGACCTGCGCGCAGCAATGCCAGGGGCTC 279
Qy 74 -----ProLysGluSerArgTrpLeuArgGluileLeuLeuCyAalaAaspGlyGlu 91
Db 280 GACGTCCCTACCGCAGTAGCGGTGGGTGGTGGCGAGGTCACCTGCATGGCCAGCAGCGT 339
Qy 92 ProTrpLeuAlaGlyArgThrValProValSerThrLeuSerGlyProGluLeuAla 111
Db 340 CCCTGGGTGTTCGCGCGCAGGTGGCGCGCGCAGCGCCCTGGAAGGCTCGGGCTTCGAC 399
Qy 112 LeuGlnLysLeuGlyLysThrProLeuGlyArgTrpLeuPheThrSerSerThrLeuThr 131
Db 400 CTGGCGCTGCTCGGCACCCCGCTGCTGGCGGAGTGTGTTTCAGCAGCAGCGCTTCGAG 459
Qy 132 ArgAaspPheIleGluile-----GlyArgAasp 140
Db 460 CGCGGGCCATCGAAGTCTCGCGTATCCGGCGCGGTCTGCCCGCGAGGTCCGCGCC 519
Qy 141 AlaGlyLeuTrpGlyArgSerArgLeuLeuSerGlyLysProLeuLeuLeuThr 160
Db 520 GAGGCTCTCTGGCGCGTCTCGTCAAGTGTTCGCGCGCGCGCTCGGGGTCTGTGGCG 579
Qy 161 GluLeuPheLeuPro 165
Db 580 GAGGTGTACCTACCG 594

RESULT 10
US-09-540-236-1889
; Sequence 1889, Application US/09540236
; Patent No. 667310
; GENERAL INFORMATION:
; APPLICANT: Gary L. Braton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709-2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 1889
; TYPE: DNA
; ORGANISM: M.catarrhalis
US-09-540-236-1889

Alignment Scores:
Pred. No.: 3,32e-12 Length: 519
Score: 174.50 Matches: 48
Percent Similarity: 49.4% Conservative: 36
Best Local Similarity: 28.2% Mismatches: 57
Query Match: 19.7% Indels: 29
DB: 3 Gaps: 7

US-10-718-311-16 (1-170) x US-09-540-236-1889 (1-519)
Qy 20 CysLysGluileProAlaLeuAasp----- 27
Db 4 TGTAACATTGTACCGCCCAATAGATGATTTCATGCCAATGGATGATATTAGACCAC 63
Qy 28 -----ProGlnLeuAaspTrpLeuLeuLeuGluAaspSerMetThrLysArgPheGlu 45
Db 288 -----ProGlnLeuAaspTrpLeuLeuLeuGluAaspSerMetThrLysArgPheGlu 45
```

```
Db 64 CTGCCTCAAAACCTTGGCGATTGTTGTTACTCAAGGTCGTTAAACGGCAGGATTAGAA 123
Qy 46 GlnGlnGly---LysThrValSerValThrMetIleArgGluGly----- 59
Db 124 GCACACTCAAAAAACCCCTAAAGGTTGAAATCATCATCAAGGCTATCAGCCACTCACC 183
Qy 60 PheValGluGlnAenGluileProGluLeuProLeuLeuProLysGluSerArgTyr 79
Db 184 TTTGCCCAAAATCATCAACTG-----GGACTGCC---CTAAATCGCCCAAAATGGCT 234
Qy 80 TrpLeuArgGluileLeuLeuCyAalaAaspGlyGluProTrpLeuAlaGlyArgThrVal 99
Db 235 TGGTAGCGAGTGTCAAGTTATACGGTATCGGACGCTTGGGTCCTGCCAAAGTAT 294
Qy 100 ValProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrPro 119
Db 295 TTTCTCTTGTTCATCCTACGTGACAGGTTTCATTAGCAGCCTAAAAAGCAGCTGGGTGCAACGCC 354
Qy 120 LeuGlyArgTyrLeuPheThrSerSer-----ThrLeuThrArgAaspPheIleGlu 136
Db 355 ATCGGATATGTTTGTTCAAAAAAGGCAACACTGCCAATTAAAGCGTCATTTTAT--- 411
Qy 137 IleGlyArgAaspAlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysPro 156
Db 412 -----CGGTATGATGTCCTCAATGGTGGCAACAGTTTATGATATCAATGGCGGTAAA 465
Qy 157 LeuLeuLeuThrGluLeuPheLeuProAla 166
Db 466 ATCTTGATGATGAGCTATTTTGGCAGCT 495

RESULT 11
US-09-596-002-41
; Sequence 41, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Patterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 41
; LENGTH: 269223
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte template ID No. 6632636 41
; PUBLICATION INFORMATION:
US-09-596-002-41

Alignment Scores:
Pred. No.: 1.09e-07 Length: 269223
Score: 171.50 Matches: 46
Percent Similarity: 53.3% Conservative: 34
Best Local Similarity: 30.7% Mismatches: 55
Query Match: 19.3% Indels: 15
DB: 3 Gaps: 6

US-10-718-311-16 (1-170) x US-09-596-002-41 (1-269223)
Qy 26 LeuAaspProGlnLeuLeuAaspTrpLeuLeuLeuAaspSerMetThrLysArgPheGlu 45
Db 16824 CTGCCTCAAAACCTTGGCGATTGTTTACTCAAGGTCGTTAAACGGCAGGATTAGAA 16883
Qy 46 GlnGlnGly---LysThrValSerValThrMetIleArgGluGly----- 59
Db 16884 GCACACTCAAAAAACCCCTAAAGGTTGAAATCATCATCAAGGCTATCAGCCACTCACC 16943
```

Qy	100	ValProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrPro	119
Db	298	TTTCCAATTCAAGATTACAAAAAAGCCGCGTATATTTTCAGCATATGGTTCTAAGCCG	355
Qy	120	LeuGlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheLeuGluLeuGlyArg	139
Db	358	ATAGTCTTTTATTTATTTCAAGAACAACACCACCTTTGTGATCGCCGGGTTATTCGTTTA	417
Qy	140	AspAlaGlyLeuTrpGlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeu	159
Db	418	CCAGNAGC---TGCAGCGACAAAGTTGTATACTTGCATGGATGTAATTTATTTGTT	474
Qy	160	ThrGluLeuPheLeuProAla	166
Db	475	CAAGAAACATTTCTTACCGGCT	495

RESULT 13  
US-09-252-991A-7361  
; Sequence 7361, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```

? FILE REFERENCE: 10/196.136
? CURRENT APPLICATION NUMBER: US/09/252,891A
? CURRENT FILING DATE: 1999-02-18
? PRIOR APPLICATION NUMBER: US 60/074,788
? PRIOR FILING DATE: 1998-02-18
? PRIOR APPLICATION NUMBER: US 60/094,190
? PRIOR FILING DATE: 1998-07-27
? NUMBER OF SEQ ID NOS: 33142
? SEQ ID NO 7361
? LENGTH: 1248
? TYPE: DNA
? ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7361

```

US-10-718-311-16 (1-170) x US-09-352-991A-7361 (1-1248)  
 Query Match: 16.5%  
 Best Local Similarity: 36.1%  
 DB: 3  
 Gaps: 1  
 Indels: 1  
 Mismatches: 1

```

5  TGGGTCGGCGAGGTCATACCTGCATCGGCCACGACCGTCTCTGGAGTGTTGCGCCGCGACGCG 84
Qy 100 ValProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrPro 119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 GCGCGCGCAGCGCCCTGGAAGGCTCGGGCTTCGACCTGGCGCTCTCGGCACCCGCTCG 124
Qy 120 LeuGlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluLeu----- 137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 CTGGCGCAGTTGCTGTTTCAGCGACAGCGCCTTCGAGCGCGCGGCCCATCGAAGTCTGCCGC 184
Qy 138 -----GlyArgAspAlaGlyLeuTyrGlyValArgGser 148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 185 TATCCGGCGGCGGTCTGCCCGCCGAGGTCGCGCGCGAGGTCTCTGGGGCCGTGCTCA 244
Qy 149 ArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 245 CGGTTTTCCCGCGCGCGCTCGGGTGCTGGTGGCGGAGGTGTACTCTACCG 295

RESULT 14
US-09-543-681A-3086/c
; Sequence 3086, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:

```

APPLICANT: GARY BRETON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

Alignment Scores:	
Pred. No.:	3,25e-06
Score:	121.00
Length:	195
Percent Similarity:	82.4%
Matches:	24
Best Local Similarity:	70.6%
Conservative:	6
Mismatches:	4
Query Match:	13.6%
Indels:	0
DB:	3
Gaps:	0

US-10-718-311-16 (1-170) x US-09-543-681A-3086 (1-195)

**Oy** IleGlyArgAspAlcLeuTrpClylYArgSerzArqLewAurSeuSgLyIysPro 156  
:::  
**Dd** 191 ATTGGCAACAAATAGGCATTGTTCCTCCCGTTTCAGCATTAACCT 132  
  
**Oy** LeuLeuLeuThrGlutPheLeuProAlaSerProLeuTyrr 170  
:::  
**Dd** 131 TTATTATAACTGAGGTGTTTTTGCTGCATCACCTGATAT 90

**RESULT 15**

```

US-09-818-780-75
; Sequence 75, Application US/09818780
; Patent No. 6677146
; GENERAL INFORMATION:
; APPLICANT: McHenry, Charles
; TITLE OF INVENTION: NOVEL THERMOPHILIC POLYMERASE III HOLOENZYME
; FILE REFERENCE: 1794.0030004
; CURRENT APPLICATION NUMBER: US/09/818,780
; CURRENT FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/192,736
; PRIOR FILING DATE: 2000-03-28
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 75
; LENGTH: 2127
; TYPE: DNA
; ORGANISM: Thermus thermophilus
US-09-818-780-75

```

Alignment Scores:		
Pred. No.:	0.875	Length: 2127
Score:	90.00	Matches: 59
Percent Similarity:	32.1%	Conservative: 22
Best Local Similarity:	23.4%	Mismatches: 77
Query Match:	10.1%	Indels: 94
DB:	3	Gaps: 10

US-10-718-311-16 (1-170) x US-09-818-780-75 (1-2127)

Qy	9	ProAlaLeuThrGlnLeuArgAla-----LeuArgTyrCysLeuGlu	22
Db	1027	CCCCGCTGGCGAGCTTAGGCCCCGCTCCTCGGGTGCCCCCTACCTACCTCCCTG	1086
Qy	23	IleProAlaLeuAsp-----ProGlnLeuLeuAspTir	33
Db	1087	GTGCCCGCGGTGGAGGTTTGGAGCGAAAGGCTTCGCCCTGCCCTGCCACAGCCCGC	1146
Qy	34	LeuLeuLeuGluAsp-----Ser	39

GenCore version 5.1.8  
 Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 30, 2006, 03:03:17 ; Search time 1170.21 Seconds  
 (without alignments)  
 2677.596 Million cell updates/sec

Title: US-10-718-311-16  
 Perfect score: 887  
 Sequence: 1 MQVWMSHPALQRLALRYC.....RLSGKPLLTELFLPASPLY 170

Scoring table: BLOSUM62  
 Xgapop 10.0, Xgapext 0.5  
 Ygapop 10.0, Ygapext 0.5  
 Fgapop 6.0, Fgapext 7.0  
 Delop 6.0, Delext 7.0

Searched: 18892170 seqs, 6143817638 residues

Total number of hits satisfying chosen parameters: 37784340

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Command line parameters:  
 -MODEL=frame+ p2n model -DRV=xlh  
 -Q-/abses/ABSWEB\_spool/US10718311/runat\_26052006\_164933\_13735/app\_query.fasta\_1  
 -DB=Published Applications NA Main -QFMT=fastap -SUFFIX=rnpbpm -MINMATCH=0.1  
 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62  
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
 -MAXLEN=2000000000 -HOST=abses05h  
 -USER=US10718311 @CGN\_1\_1 2326 @runat\_26052006\_164933\_13735 -NCPUP=6 -ICPU=3  
 -NO\_MMAP -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV\_TIMEOUT=120  
 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA\_Main:

1:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US08_PUBCOMB.seq.*
3:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
4:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
5:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
6:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
7:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
8:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
9:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
10:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
11:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
12:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
13:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US10H_PUBCOMB.seq.*
14:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US10I_PUBCOMB.seq.*
15:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US10J_PUBCOMB.seq.*
16:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US10K_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	887	100.0	684	3	US-09-855-341-7 Sequence 7, Appli

2	887	100.0	684	3	US-09-896-866B-16	Sequence 16, Appl
3	887	100.0	684	7	US-10-359-369-41	Sequence 41, Appl
4	887	100.0	684	7	US-10-699-050-16	Sequence 16, Appl
5	887	100.0	684	8	US-10-718-311-7	Sequence 7, Appli
6	887	100.0	684	9	US-10-462-162-33	Sequence 33, Appl
7	862	97.2	3452	9	US-10-462-162-47	Sequence 47, Appl
8	854	96.3	495	3	US-09-855-341-3	Sequence 3, Appli
9	854	96.3	495	3	US-09-896-866B-12	Sequence 12, Appl
10	854	96.3	495	7	US-10-359-369-37	Sequence 37, Appl
11	854	96.3	495	8	US-10-699-050-12	Sequence 12, Appl
12	854	96.3	495	8	US-10-718-311-3	Sequence 3, Appli
13	854	96.3	498	9	US-10-462-162-29	Sequence 29, Appl
14	851	95.9	1971	9	US-10-462-162-54	Sequence 54, Appl
15	811	91.4	1207	10	US-10-450-763-26139	Sequence 26139, A
16	171.5	19.3	269223	8	US-10-672-787-41	Sequence 41, Appl
17	97.5	11.0	475	8	US-10-325-026A-113	Sequence 113, App
18	91	10.3	606	4	US-09-925-065A-789463	Sequence 789463,
19	91	10.3	606	5	US-09-925-065A-789463	Sequence 789463,
20	90	10.1	2127	9	US-10-746-167-75	Sequence 75, Appl
21	90	10.1	9133	10	US-10-915-740A-277	Sequence 277, App
22	90	10.1	2242716	10	US-10-915-740A-1068	Sequence 1068, Ap
23	88.5	10.0	2892	16	US-11-183-136-15	Sequence 15, Appl
24	88	9.9	42811	11	US-10-330-773-925	Sequence 925, App
25	87	9.8	59589	6	US-10-087-192-1315	Sequence 1315, Ap
26	86	9.7	522	10	US-10-467-657-2855	Sequence 2855, Ap
27	85.5	9.6	2907	7	US-10-108-260A-1067	Sequence 1067, Ap
28	85	9.6	888	10	US-10-617-320-1721	Sequence 1721, Ap
29	84.5	9.5	1780	10	US-10-487-078-92	Sequence 92, Appl
30	84.5	9.5	11331	10	US-10-461-862-101	Sequence 101, App
31	84	9.5	25001	7	US-10-017-161-2063	Sequence 2063, Ap
32	84	9.5	25001	7	US-10-292-798-1709	Sequence 1709, Ap
33	83.5	9.4	1514	8	US-10-424-599-27291	Sequence 27291, A
34	83.5	9.4	1565	10	US-10-750-185-33285	Sequence 33285, A
35	83.5	9.4	1565	10	US-10-750-623-33285	Sequence 33285, A
36	83.5	9.4	10710	8	US-10-672-396-4	Sequence 4, Appli
37	83.5	9.4	196200	15	US-11-121-086-9	Sequence 9, Appli
38	83.5	9.4	199321	15	US-11-121-086-10	Sequence 10, Appl
39	83	9.4	608	4	US-09-925-065A-772942	Sequence 772942,
40	83	9.4	608	5	US-09-925-065A-772942	Sequence 772942,
41	83	9.4	1140	3	US-09-928-175-22	Sequence 22, Appl
42	83	9.4	2314	3	US-09-928-175-19	Sequence 19, Appl
43	83	9.4	2337	8	US-10-437-963-39118	Sequence 39118, A
44	83	9.4	2539	6	US-10-229-735-4	Sequence 4, Appli
45	82.5	9.3	689	13	US-11-097-143-18275	Sequence 18275, A

## ALIGNMENTS

RESULT 1  
 US-09-855-341-7  
 ; Sequence 7, Application US/09855341  
 ; Patent No. US20020002715A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VIITANEN, PAUL V.  
 ; APPLICANT: MEYER, KNUD  
 ; APPLICANT: VAN DYK, DREW  
 ; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID  
 ; TITLE OF INVENTION: IN GREEN PLANTS  
 ; FILE REFERENCE: BC1015 US NA  
 ; CURRENT APPLICATION NUMBER: US/09/855,341  
 ; CURRENT FILING DATE: 2001-05-15  
 ; NUMBER OF SEQ ID NOS: 16  
 ; SOFTWARE: MICROSOFT OFFICE 97  
 ; SEQ ID NO 7  
 ; LENGTH: 684  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence:synthetic CPL  
 US-09-855-341-7

Alignment Scores:  
 Pred. No.: 9.84e-106 Length: 684



Score: 887.00 Matches: 170  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 100.0% Indels: 0  
 DB: 3 Gaps: 0

US-10-718-311-16 (1-170) x US-09-855-341-7 (1-684)

Qy 1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20  
 Db 172 ATGAGGTGTGGCATATGTACACCCCGGTTAAGCAACTCGGTGCGCTATTGT 231  
 Qy 21 LysGluIleProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40  
 Db 232 AAAGAGATCCCTGCCCTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGATTCCATG 291  
 Qy 41 ThrLysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPhe 60  
 Db 292 ACNAAACGTTTGAACACAGGAGAAACCGGTAGCGTGACGATGATCCGCGAAGGTTT 351  
 Qy 61 ValGluGlnAnGluIleProGluLeuProLeuLeuProLysGluSerArgTyrTrp 80  
 Db 352 GTCGAGCAGAAATGAAATCCCGAAGAACTGCGCTGCTGCGGAAAGAGTCTCGTTACTGG 411  
 Qy 81 LeuArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100  
 Db 412 TTACGTGAAATTTTGTATGTGCCGATGGTGAACCGTGGCTTGGCGGTGCTGCTGCTG 471  
 Qy 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120  
 Db 472 CCTGTGTCAACGTTAAGCGGCGGAGCTGGCTTTACAAAATTTGGGTAAACCGCGTTA 531  
 Qy 121 GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAsp 140  
 Db 532 GGACGCTATCTGTTCCACATCATCGACATTAACCGGGACTTTATTGAGATAGCGCGTGAT 591  
 Qy 141 AlaGlyLeuTrpGlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThr 160  
 Db 592 GCCGGGCTGTGGGGCGAGCTTCCCGCTTGGATTAAGCGGTAAACCGCTGTGCTAACA 651

Qy 161 GluLeuPheLeuProAlaSerProLeuTyr 170  
 Db 652 GAACGTGTTTACCGCGGTACCGGTTGTAC 681

RESULT 2  
 US-09-896-866B-16  
 ; Sequence 16, Application US/09896866B  
 ; Patent No. US20020151002A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Flint, Dennis  
 ; APPLICANT: Meyer, Knut  
 ; APPLICANT: Viitanen, Paul  
 ; TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugates  
 ; FILE REFERENCE: BC1034 US NA  
 ; CURRENT APPLICATION NUMBER: US/09/896,866B  
 ; PRIOR FILING DATE: 2001-08-29  
 ; PRIOR FILING DATE: 2000-07-07  
 ; NUMBER OF SEQ ID NOS: 19  
 ; SOFTWARE: Microsoft Office 97  
 ; SEQ ID NO 16  
 ; LENGTH: 684  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; OTHER INFORMATION: open reading frame of the chloroplast-targeted CPL fusion protein  
 US-09-896-866B-16

Alignment Scores: 9.84e-106 Length: 684  
 Pred. No.: 887.00 Matches: 170  
 Score:  
 Percent Similarity: 100.0% Conservative: 0

Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 100.0% Indels: 0  
 DB: 3 Gaps: 0

US-10-718-311-16 (1-170) x US-09-896-866B-16 (1-684)

Qy 1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20  
 Db 172 ATGAGGTGTGGCATATGTACACCCCGGTTAAGCAACTCGGTGCGCTATTGT 231  
 Qy 21 LysGluIleProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40  
 Db 232 AAAGAGATCCCTGCCCTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGATTCCATG 291  
 Qy 41 ThrLysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPhe 60  
 Db 292 ACNAAACGTTTGAACACAGGAGAAACCGGTAGCGTGACGATGATCCGCGAAGGTTT 351  
 Qy 61 ValGluGlnAnGluIleProGluLeuProLeuLeuProLysGluSerArgTyrTrp 80  
 Db 352 GTCGAGCAGAAATGAAATCCCGAAGAACTGCGCTGCTGCGGAAAGAGTCTCGTTACTGG 411  
 Qy 81 LeuArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100  
 Db 412 TTACGTGAAATTTTGTATGTGCCGATGGTGAACCGTGGCTTGGCGGTGCTGCTGCTG 471  
 Qy 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120  
 Db 472 CCTGTGTCAACGTTAAGCGGCGGAGCTGGCTTTACAAAATTTGGGTAAACCGCGTTA 531  
 Qy 121 GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAsp 140  
 Db 532 GGACGCTATCTGTTCCACATCATCGACATTAACCGGGACTTTATTGAGATAGCGCGTGAT 591  
 Qy 141 AlaGlyLeuTrpGlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThr 160  
 Db 592 GCCGGGCTGTGGGGCGAGCTTCCCGCTTGGATTAAGCGGTAAACCGCTGTGCTAACA 651

Qy 161 GluLeuPheLeuProAlaSerProLeuTyr 170  
 Db 652 GAACGTGTTTACCGCGGTACCGGTTGTAC 681

RESULT 3  
 US-10-359-369-41  
 ; Sequence 41, Application US/10359369  
 ; Publication No. US20030215927A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: E. I. duPont de Nemours and Company, Inc.  
 ; APPLICANT: Viitanen, Paul  
 ; APPLICANT: Meyer, Knut  
 ; APPLICANT: Van Dyk, Drew  
 ; TITLE OF INVENTION: UDP-Glucosyltransferases  
 ; FILE REFERENCE: CL1821 US NA  
 ; CURRENT APPLICATION NUMBER: US/10/359,369  
 ; CURRENT FILING DATE: 2003-02-06  
 ; PRIOR FILING DATE: 2002-02-07  
 ; NUMBER OF SEQ ID NOS: 46  
 ; SOFTWARE: Microsoft Office 97  
 ; SEQ ID NO 41  
 ; LENGTH: 684  
 ; TYPE: DNA  
 ; ORGANISM: Escherichia coli  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)...(684)  
 US-10-359-369-41

Alignment Scores: 9.84e-106 Length: 684  
 Pred. No.: 887.00 Matches: 170  
 Score:  
 Percent Similarity: 100.0% Conservative: 0



Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 7 Gaps: 0

US-10-718-311-16 (1-170) x US-10-359-369-41 (1-684)

QY 1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20  
DB 172 ATGCAGGTGTGGCATATGTACACCCCGGTTAAACCAACTGCGTGGCTATTGT 231  
QY 21 LysGluileProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuLeuAspSerMet 40  
DB 232 AAAGAGATCCCTGCGCTGGATCCGCAACTGCTGACTGGTGTGTGCTGAGGATCCATG 291  
QY 41 ThrLysArgPheGluGlnGlnGlyLysThrValSerValThrMetileArgGluGlyPhe 60  
DB 292 ACAAAACGTTTTGAACAGCAGGGAACAAACGGTAAGCGTGACGATGATCCGCAAGGTTT 351  
QY 61 ValGluGlnAsnGluileProGluGluLeuProLeuLeuProLysGluSerArgTyrTrp 80  
DB 352 GTCCGAGCAGATGAATCCCGAAGAACTGCCGCTGCTGCCGAAAGAGTCTCGTTACTGG 411  
QY 81 LeuArgGluileLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100  
DB 412 TTACGTGAAATTTGTATGTGCCGATGGTGAACCGTGGCTTGCCTGCGTTCGTCGTT 471  
QY 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120  
DB 472 CCGTGTCAACGTTAAGCGGGCGGAGCTGGGTTACAAAAATTTGGTTAAACGCGTTA 531  
QY 121 GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheileGluileGlyArgAsp 140  
DB 532 GCAGCGTATCTGTTACATCATCGACATTAACCCGGGACTTTATTGAGATAGCCGCTGAT 591  
QY 141 AlaGlyLeuTrpGlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThr 160  
DB 592 GCGGGCTGTGGGGCGAGTTCGCCCTGCGATTAAAGCGGTAAACCCGCTGTTGCTAAC 651  
QY 161 GluLeuPheLeuProAlaSerProLeuTyr 170  
DB 652 GAACGTGTTTTACCGGCGTCACCGTTGTAC 681

RESULT 4

US-10-699-050-16  
; Sequence 16, Application US/10699050  
; Publication No. US20040142437A1  
; GENERAL INFORMATION:  
; APPLICANT: Flint, Dennis  
; APPLICANT: Meyer, Knut  
; APPLICANT: Viitanen, Paul  
; TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugates  
; FILE REFERENCE: BC1034 US NA  
; CURRENT APPLICATION NUMBER: US/10/699,050  
; CURRENT FILING DATE: 2003-10-30  
; PRIOR APPLICATION NUMBER: US/09/896,866B  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: 60/216,615  
; PRIOR FILING DATE: 2000-07-07  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 16  
; LENGTH: 684  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; NAME/KEY: misc feature  
; OTHER INFORMATION: open reading frame of the chloroplast-targeted CPL fusion protein

US-10-699-050-16

Alignment Scores:  
Pred. No.: 9,848-106 Length: 684  
Score: 887.00 Matches: 170

Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 8 Gaps: 0

US-10-718-311-16 (1-170) x US-10-699-050-16 (1-684)

QY 1 MetGlnValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20  
DB 172 ATGCAGGTGTGGCATATGTACACCCCGGTTAAACCAACTGCGTGGCTATTGT 231  
QY 21 LysGluileProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuLeuAspSerMet 40  
DB 232 AAAGAGATCCCTGCGCTGGATCCGCAACTGCTGACTGGTGTGTGCTGAGGATCCATG 291  
QY 41 ThrLysArgPheGluGlnGlnGlyLysThrValSerValThrMetileArgGluGlyPhe 60  
DB 292 ACAAAACGTTTTGAACAGCAGGGAACAAACGGTAAGCGTGACGATGATCCGCAAGGTTT 351  
QY 61 ValGluGlnAsnGluileProGluGluLeuProLeuLeuProLysGluSerArgTyrTrp 80  
DB 352 GTCCGAGCAGATGAATCCCGAAGAACTGCCGCTGCTGCCGAAAGAGTCTCGTTACTGG 411  
QY 81 LeuArgGluileLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100  
DB 412 TTACGTGAAATTTGTATGTGCCGATGGTGAACCGTGGCTTGCCTGCGTTCGTCGTT 471  
QY 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeu 120  
DB 472 CCGTGTCAACGTTAAGCGGGCGGAGCTGGGTTACAAAAATTTGGGTTAAACGCGTTA 531  
QY 121 GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheileGluileGlyArgAsp 140  
DB 532 GCAGCGTATCTGTTACATCATCGACATTAACCCGGGACTTTATTGAGATAGCCGCTGAT 591  
QY 141 AlaGlyLeuTrpGlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThr 160  
DB 592 GCGGGCTGTGGGGCGAGTTCGCCCTGCGATTAAAGCGGTAAACCCGCTGTTGCTAAC 651  
QY 161 GluLeuPheLeuProAlaSerProLeuTyr 170  
DB 652 GAACGTGTTTTACCGGCGTCACCGTTGTAC 681

RESULT 5

US-10-718-311-7  
; Sequence 7, Application US/10718311  
; Publication No. US20040143867A1  
; GENERAL INFORMATION:  
; APPLICANT: Viitanen, Paul V.  
; APPLICANT: MEYER, KNUT  
; APPLICANT: VAN DYK, DREW  
; TITLE OF INVENTION: HIGH LEVEL PRODUCTION OF P-HYDROXYBENZOIC ACID  
; FILE REFERENCE: BC1015 US NA  
; CURRENT APPLICATION NUMBER: US/10/718,311  
; CURRENT FILING DATE: 2003-11-20  
; PRIOR APPLICATION NUMBER: US/09/855,341  
; PRIOR FILING DATE: 2001-05-15  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: MICROSOFT OFFICE 97  
; SEQ ID NO 7  
; LENGTH: 684  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:synthetic CPL

US-10-718-311-7

Alignment Scores:  
Pred. No.: 9,848-106 Length: 684  
Score: 887.00 Matches: 170  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0

```

Query Match: 100.0% Indels: 0
DB: 8 Gaps: 0

US-10-718-311-16 (1-170) x US-10-718-311-7 (1-684)

Qy 1 MetGlnValTrrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20
Db 172 ATGAGGTGGCATAATGTCACCCCGGTTAACGCACTGCGTGGCTCGCTATTGT 231
Qy 21 LysGluileProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40
Db 232 AAAGAGATCCCTGCGCTGGATCGCAACTGCTCGACTGGCTGTTGCTGGAGATTCCATG 291
Qy 41 ThrIysArgPheGluGlnGlnGlyThrValSerValThrMetIleArgGluGlyPhe 60
Db 292 ACAAAACGTTTTTGAACACGAGGGAACACGGTAAGCGTGACGATGATCCGGAAGGGTTT 351
Qy 61 ValGluGlnAenGluileProGluGluLeuProLeuProLysGluSerArgTyrTrp 80
Db 352 GTCGAGCAGAATGAATCCCGAAGAACTGCGCGTGTCTGCGAAGAGTCTCGTTACTGG 411
Qy 81 LeuArgGluileLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100
Db 412 TTACGTGAAATTTTGTATGTGCGCATGTTGAACCGTGGCTTGCCTGCTGACCGTGT 471
Qy 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyThrProLeu 120
Db 472 CCTGTGTCAACGTTAAGCGGGCGGAGCTGCGGTTACAAAAATTTGGGTAAACCGCGTTA 531
Qy 121 GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluileGlyArgAsp 140
Db 532 GGACGCTATCTGTTACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCGGTGAT 591
Qy 141 AlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 160
Db 592 GCCGGCTGTGGGGCGACGTTCCCGCTCGATTAAAGCGGTAACCGCTGTGCTAACA 651

RESULT 6
US-10-462-162-33
; Sequence 33, Application US/10462162
; Publication No. US20040261147A1
; GENERAL INFORMATION:
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; APPLICANT: Flint, Dennis
; TITLE OF INVENTION: High Level Production of Arbutin in Green Plants and Microbes
; FILE REFERENCE: CL 2155 US NA
; CURRENT APPLICATION NUMBER: US/10/462,162
; CURRENT FILING DATE: 2003-06-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 33
; LENGTH: 684
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-462-162-33

Alignment Scores:
Pred. No.: 9.84e-106 Length: 684
Score: 887.00 Matches: 170
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 9 Gaps: 0

US-10-718-311-16 (1-170) x US-10-462-162-33 (1-684)

Qy 1 MetGlnValTrrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCys 20

```

```

Db 172 ATGAGGTGGCATAATGTCACCCCGGTTAACGCACTGCGTGGCTCGCTATTGT 231
Qy 21 LysGluileProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMet 40
Db 232 AAAGAGATCCCTGCGCTGGATCGCAACTGCTCGACTGGCTGTTGCTGGAGATTCCATG 291
Qy 41 ThrIysArgPheGluGlnGlnGlyThrValSerValThrMetIleArgGluGlyPhe 60
Db 292 ACAAAACGTTTTTGAACACGAGGGAACACGGTAAGCGTGACGATGATCCGGAAGGGTTT 351
Qy 61 ValGluGlnAenGluileProGluGluLeuProLeuProLysGluSerArgTyrTrp 80
Db 352 GTCGAGCAGAATGAATCCCGAAGAACTGCGCGTGTCTGCGAAGAGTCTCGTTACTGG 411
Qy 81 LeuArgGluileLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValVal 100
Db 412 TTACGTGAAATTTTGTATGTGCGCATGTTGAACCGTGGCTTGCCTGCTGACCGTGT 471
Qy 101 ProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyThrProLeu 120
Db 472 CCTGTGTCAACGTTAAGCGGGCGGAGCTGCGGTTACAAAAATTTGGGTAAACCGCGTTA 531
Qy 121 GlyArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGluileGlyArgAsp 140
Db 532 GGACGCTATCTGTTACATCATCGACATTAACCCGGGACTTTATTGAGATAGGCGGTGAT 591
Qy 141 AlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThr 160
Db 592 GCCGGCTGTGGGGCGACGTTCCCGCTCGATTAAAGCGGTAACCGCTGTGCTAACA 651

RESULT 7
US-10-462-162-47
; Sequence 47, Application US/10462162
; Publication No. US20040261147A1
; GENERAL INFORMATION:
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; APPLICANT: Flint, Dennis
; TITLE OF INVENTION: High Level Production of Arbutin in Green Plants and Microbes
; FILE REFERENCE: CL 2155 US NA
; CURRENT APPLICATION NUMBER: US/10/462,162
; CURRENT FILING DATE: 2003-06-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 47
; LENGTH: 3452
; TYPE: DNA
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Nucleotide sequence of nucleic acid fragment encoding CPL (from
; OTHER INFORMATION: Escherichia coli), pRBA 1-hydroxylase (from Candida
; OTHER INFORMATION: parapsilosis), and UGT72B1 (from Arabidopsis thaliana).
US-10-462-162-47

Alignment Scores:
Pred. No.: 1.78e-101 Length: 3452
Score: 862.00 Matches: 166
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 97.2% Indels: 0
DB: 9 Gaps: 0

US-10-718-311-16 (1-170) x US-10-462-162-47 (1-3452)

Qy 5 HisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLeuGluilePro 24
Db 1 CATATGTACACCCCGCGTTAAACGCAACTGCGTGGCTGCGCTATTGTAAAGAGATCCCT 60

```

Qy	46	GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu	65
Db	121	CAGCAGGGAAAAACGGTAAAGCGTGACGATATCCGCGAAGGGCTTTGTCTCAGCAGAAATGAA	180
Qy	66	IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu	85
Db	181	ATCCCCGAAGAACATCGCCCTGCTGCCGAAGAGTCTCGTTACTGTGTACGTGAAATTTTGT	240
Qy	86	LeuCyAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu	105
Db	241	TTATGTCCGATGGTGAACCGTGGCTTCCCGGTCTACCGTCTCTCTGTGTCAACGTTA	300
Qy	106	SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe	125
Db	301	AGCGGGCCGAGCTGGCGTTACAAAAATTGGGTAAAAACGCCGTTAGGACGCTATCTGTTC	360
Qy	126	ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly	145
Db	361	ACATCATCGACATTAACCCGGGACTTTATTGAGATAGCGCGGTGATGCCGGGCTGTGGGG	420
Qy	146	ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro	165
Db	421	CGAGCTTCCCCTCCGATTAAGCGGTAAAACCGCTGTGTCTAACAGACTGTTTATACCG	480
Qy	166	AlaSerProLeuTyr	170
Db	481	GCGTCAACCGTTGTAC	495

RESULT 9  
US-09-896-866B-12  
; Sequence 12, Application US/09896866B  
; Patent No. US20020151002A1  
; GENERAL INFORMATION:  
; APPLICANT: Flint, Dennis  
; APPLICANT: Meyer, Knut  
; APPLICANT: Viitanen, Paul  
; TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugate  
; TITLE OF INVENTION: Benzoyl Acid Glucosides  
; FILE REFERENCE: BC1034 US NA  
; CURRENT APPLICATION NUMBER: US/09/896,866B  
; CURRENT FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: 60/216,615  
; PRIOR FILING DATE: 2000-07-07  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 12  
; LENGTH: 495  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
US-09-896-866B-12

```
Alignment Scores:
Pred. No.:      1.35e-101
Score:          854.00
Percent Similarity: 100.0%
Best Local Similarity: 100.0%
Query Match:    96.3%
DB:             3
                6
                3
                0
                0
Length:         1955
Matches:        1650
Conservative:   0
Mismatch:       0
Indels:         0
Gaps:           0
US-10-718-311-16 (1-170) x US-09-896-866B-12 (1-495)
```

Qy	6	MetSerHisProAlaLeuThrGlnLeuArgAlaIeuArgTyrCysLysGluIleProAla	25
Db	1	ATGTCAACCCCGCGTTAACGCAACTGCGTGGCTGGCTATTGTAAAGAGATCCCTGCC	60
Qy	26	LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu	45
Db	61	CTGGATCCGCAACTGCTCGACTGGCTGTGTCTGGAGGATTCCATGACAAACCGTTTGA	120
Qy	46	GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu	65
Db	121	CAGCAGGAGAAAACGCTTAAGCGTACCATGATCCGCAAGGTTTGTTCAGCGAGATGAA	180

```

Qy 66 IleProGluLeuProLeuProLeuProLysGluSerArgTyrTrpLeuArgGluLeu 85
Db 181 ATCCCGAAGAACTGCGCTGCTCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTTG 240
Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 105
Db 241 TTATGTCCGATGGTAACCGTGGCTTGCCGGTCTGTAACCGTTTCTGTGTCAACGTTA 300
Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
Db 301 AGCGGCGGAGCTGGCTTACAAAATTTGGTAAACCGCGTTAGGACGCTACTCTGTTTC 360
Qy 126 ThrSerThrLeuThrArgAspPheIleGluLeuGlyArgAlaGlyLeuTrpGly 145
Db 361 ACATCATCGACATTAACCCGGGACTTTATTAGATAGCGCGTGTATGCGGGCTGTGGGG 420
Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 165
Db 421 CGAGCTTCCCGCCCTGCGATTAAAGCGGTAAACCGCTGTTGCTTAAACAGAACTGTTTACCG 480
Qy 166 AlaSerProLeuTyr 170
Db 481 CGGTACCGTTGTAC 495

```

RESULT 10

US-10-359-369-37

; Sequence 37, Application US/10359369

; Publication No. US20030215927A1

; GENERAL INFORMATION:

; APPLICANT: E. I. duPont de Nemours and Company, Inc.

; APPLICANT: Viitanen, Paul

; APPLICANT: Meyer, Knut

; APPLICANT: Van Dyk, Drew

; TITLE OF INVENTION: UDP-Glucosyltransferases

; FILE REFERENCE: C11821 US NA

; CURRENT APPLICATION NUMBER: US/10/359,369

; CURRENT FILING DATE: 2003-02-06

; PRIOR APPLICATION NUMBER: 60/355,511

; PRIOR FILING DATE: 2002-02-07

; NUMBER OF SEQ ID NOS: 46

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 37

; LENGTH: 495

; TYPE: DNA

; ORGANISM: Escherichia coli

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(495)

US-10-359-369-37

Alignment Scores:

Pred. No.:	1.35e-101	Length:	495
Score:	854.00	Matches:	165
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	96.3%	Indels:	0
DB:	7	Gaps:	0

US-10-718-311-16 (1-170) x US-10-359-369-37 (1-495)

```

Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 25
Db 1 ATGTACACCCCGCGTTAAACGCAACTGCGTGGCTGCTATTGTAAAGAGATCCCTGCC 60
Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAAACGTTTGA 120
Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAsnGlu 65
Db 121 CAGCAGGGAAAAACGGTAAGCGTGACGATGATCCGCGAAGGGTTGTTCGACAGAA 180

```

```

Qy 66 IleProGluLeuProLeuProLysGluSerArgTyrTrpLeuArgGluLeu 85
Db 181 ATCCCGAAGAACTGCGCTGCTCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTTG 240
Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeu 105
Db 241 TTATGTCCGATGGTAACCGTGGCTTGCCGGTCTGTAACCGTTTCTGTGTCAACGTTA 300
Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
Db 301 AGCGGCGGAGCTGGCTTACAAAATTTGGTAAACCGCGTTAGGACGCTACTCTGTTTC 360
Qy 126 ThrSerThrLeuThrArgAspPheIleGluLeuGlyArgAlaGlyLeuTrpGly 145
Db 361 ACATCATCGACATTAACCCGGGACTTTATTAGATAGCGCGTGTATGCGGGCTGTGGGG 420
Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 165
Db 421 CGAGCTTCCCGCCCTGCGATTAAAGCGGTAAACCGCTGTTGCTTAAACAGAACTGTTTACCG 480
Qy 166 AlaSerProLeuTyr 170
Db 481 CGGTACCGTTGTAC 495

```

RESULT 11

US-10-699-050-12

; Sequence 12, Application US/10699050

; Publication No. US20040142437A1

; GENERAL INFORMATION:

; APPLICANT: Flint, Dennis

; APPLICANT: Meyer, Knut

; APPLICANT: Viitanen, Paul

; TITLE OF INVENTION: Sinapoylglucose:Malate Sinapoyltransferase Form Malate Conjugates

; TITLE OF INVENTION: Benzoic Acid Glucosides

; FILE REFERENCE: BC1034 US NA

; CURRENT APPLICATION NUMBER: US/10/699,050

; CURRENT FILING DATE: 2003-10-30

; PRIOR APPLICATION NUMBER: US/09/896,866B

; PRIOR FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: 60/216,615

; PRIOR FILING DATE: 2000-07-07

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 12

; LENGTH: 495

; TYPE: DNA

; ORGANISM: Escherichia coli

US-10-699-050-12

Alignment Scores:

Pred. No.:	1.35e-101	Length:	495
Score:	854.00	Matches:	165
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	96.3%	Indels:	0
DB:	8	Gaps:	0

US-10-718-311-16 (1-170) x US-10-699-050-12 (1-495)

```

Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 25
Db 1 ATGTACACCCCGCGTTAAACGCAACTGCGTGGCTGCTATTGTAAAGAGATCCCTGCC 60
Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
Db 61 CTGGATCCGCAACTGCTCGACTGGCTGTGCTGGAGGATTCATGACAAAAACGTTTGA 120
Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnAsnGlu 65
Db 121 CAGCAGGGAAAAACGGTAAGCGTGACGATGATCCGCGAAGGGTTGTTCGACAGAA 180
Qy 66 IleProGluLeuProLeuProLysGluSerArgTyrTrpLeuArgGluLeu 85

```

Qy		86	LeuCysAlaAspGlyGluProTrrIpeuAlaGlyArgThrValValProValSerThrLeu	105
Dd		241	TTATGTCCGATGTGAACCGTGCTGCCGTTCGGTTCCTGTGTCAACGTTA	300
Qy		106	SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrPrtLeuGlyVargTyrlLeuPhe	125
Dd		301	AGCGGGCCGAGCTGGCGTTACAAAATTGGGTAAAAAGCCGTTAGGACGCATCTCTTTC	360
Qy		126	ThrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGly	145
Dd		361	ACATCATCGACATTAACCCGGGACTTTATTGAGATAGCCCGGTGATCCCGGGTGTGGGG	420
Qy		146	ArgArgSerArgLeuAsgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuPro	165
Dd		421	CGAGCTTCCC GCCCTTCGGATTAAAGCGGTAAAACGCTGTGTCTAACAGAAGCTGTTTTACC	480
Qy		166	AlaSerProLeuTyr	170
Dd		481	GCGTCAACGTTGTAC	495

RESULT 13

```

US-10-462-162-29
; Sequence 29, Application US/10462162
; Publication No. US20040261147A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Meyer, Knut
; APPLICANT: Viitanen, Paul
; APPLICANT: Flint, Dennis
; TITLE OF INVENTION: High Level Production of Arbutin in Green Plants and Microbes
; FILE REFERENCE: CL 2155 US NA
; CURRENT APPLICATION NUMBER: US/10/462.162
; CURRENT FILING DATE: 2003-06-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 29
; LENGTH: 498
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-462-162-29

```

Alignment Scores:

Alignment Scores:		
Pred. No.:	1.36E-101	Length: 498
Score:	854.00	Matches: 165
Percent Similarity:	100.0%	Conservative: 0
Best Local Similarity:	100.0%	Mismatches: 0
Query Match:	96.3%	Indels: 0
DB:	9	Gaps: 0

US-10-718-311-16 (1-170) x US-10-462-162-29 (1-498)

Qy	6	MetSerHisProAlaLeuThrGlnLeuAargAlaLeuArgTyrCysLysGluIleProAla	25
Db	1	ATGTCAACACCCCGGTTTAAGCAACTGCGTGCCTATTGTAAAGAGATCCCTGCC	60
Qy	26	LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu	45
Db	61	CTGGATCCGCAACTGCCTCGACTGGCTGTGTCTGAGGAGATTCATGACAAAACGTTTGTGAA	120
Qy	46	GlnGlnGlyThrValSerValThrMetIleAargGluGlyPheValGluGlnAsnGlu	65
Db	121	CACAGGGAAAAACGGTAAGCGTGACGATATCCGCGAAGGGTTGTCTGACGACAGATGAA	180
Qy	66	IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu	85
Db	181	ATCCCCGAAGAACTGCCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAATTTTG	240
Qy	86	LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu	105
Db	241	TTATGTGCCCATGCTGAACCGTGGCTTGGCCGTCGTACCGTCGTTCGTGTGTCACAGTTA	300
Qy	106	SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe	125

Db 301 AGCGGGCGGAGCTGGCGTTACAAAAATTGGTAAACCGCGTTAGGACGCTATCTGTTTC 360  
Qy 126 ThrSerThrLeuThrArgAspPheileGluileGlyArgAspAlaGlyLeuTrpGly 145  
Db 361 ACATCATCGACATTAACCCGGGACTTTATTGAGATAGCCGCTGATGCCGGCTGTGGGG 420  
Qy 146 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeuPro 165  
Db 421 CGAGCTTCCCGCCCTGCGATTAAAGCGGTAAACCGCTGTTGCTTAAACAGAACTGTTTTACCG 480  
Qy 166 AlaSerProLeuTyr 170  
Db 481 CCGTCACCGTTGTATC 498  
RESULT 14  
US-10-462-162-54  
; Sequence 54, Application US/10462162  
; Publication No. US20040261147A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Meyer, Knut  
; APPLICANT: Viitanen, Paul  
; APPLICANT: Flint, Dennis  
; TITLE OF INVENTION: High Level Production of Arbutin in Green Plants and Microbes  
; FILE REFERENCE: CL 2155 US NA  
; CURRENT APPLICATION NUMBER: US/10/462,162  
; CURRENT FILING DATE: 2003-06-16  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 54  
; LENGTH: 1971  
; TYPE: DNA  
; ORGANISM: artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Nucleic acid sequence of a nucleic acid fragment inserted into  
; OTHER INFORMATION: expression vector pET29a encoding CPL (from Escherichia coli) and  
; OTHER INFORMATION: pHEA 1-hydroxylase (from Cadida parapsilosis).  
US-10-462-162-54  
Alignment Scores:  
Pred. No.: 2,27e-100 Length: 1971  
Score: 851.00 Matches: 165  
Percent Similarity: 99.4% Conservative: 0  
Best Local Similarity: 99.4% Mismatches: 1  
Query Match: 9 Indels: 0  
DB: 9 Gaps: 0  
US-10-718-311-16 (1-170) x US-10-462-162-54 (1-1971)  
Qy 5 HisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluilePro 24  
Db 1 CATATGTACACCCCGCGTTAAACGCACTGGCTGGCTGCTATTTTAAAGAGATCCCT 60  
Qy 25 AlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPhe 44  
Db 61 GCCCTGGATCCGCAACTGCTGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTT 120  
Qy 45 GluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsn 64  
Db 121 GAACAGCAGGGAACAAACGTAAGCGTGACGATGATCCCGAAGGTTTGTGAGCAGAA 180  
Qy 65 GluileProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluile 84  
Db 181 GAAATCCCGAAGAACTCCGCTGCTGCGCAAGAGTCTCGTTACTGTTACGTGAAAT 240  
Qy 85 LeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThr 104  
Db 241 TTGTTATGTCGATGGTGAACCCGTTGGCTTGGCGGTGTTACGTTCTGTTGTCACG 300  
Qy 105 LeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeu 124  
Db 301 TTAAGCGCGCGAGCTGGCTTACAAAAATTGGGTAAACCGCGCTATAGGACGCTATCTG 360

Qy 125 PheThrSerSerThrLeuThrArgAspPheileGluileGlyArgAspAlaGlyLeuTrp 144  
Db 361 TTACATCATCGACATTAACCCGGGACTTTATTGAGATAGCCGCTGATGCCGGCTGTGG 420  
Qy 145 GlyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuThrGluLeuPheLeu 164  
Db 421 GGGCGAGCTTCCCGCCCTGCGATTAAAGCGGTAAACCGCTGTTGCTTAAACAGAACTGTTTTTA 480  
Qy 165 ProAlaSerProLeuTyr 170  
Db 481 CCGCGCTCACCGTTGTATC 498  
RESULT 15  
US-10-450-763-26139  
; Sequence 26139, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 790CIP3/US  
; CURRENT APPLICATION NUMBER: US/10/450,763  
; CURRENT FILING DATE: 2003-06-11  
; PRIOR APPLICATION NUMBER: PCT/US01/08631  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 26139  
; LENGTH: 1207  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIMILAR  
; LOCATION: (893)..(925)  
; OTHER INFORMATION: 81% homologous to Escherichia coli 4-hydroxybenzoate  
; OTHER INFORMATION: synthetase, accession number X57434, Smith-Waterman Score=49.  
US-10-450-763-26139  
Alignment Scores:  
Pred. No.: 2,05e-95 Length: 1207  
Score: 811.00 Matches: 163  
Percent Similarity: 98.2% Conservative: 1  
Best Local Similarity: 97.6% Mismatches: 1  
Query Match: 91.4% Indels: 2  
DB: 10 Gaps: 0  
US-10-718-311-16 (1-170) x US-10-450-763-26139 (1-1207)  
Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluileProAla 25  
Db 652 ATGTACACCCCGCGTTAAACGCACTGGCTGGCTGCTATTTGTAAGAGATCCCTGCC 711  
Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45  
Db 712 CTGGATCCGCAACTGCTGCGACTGGCTGTGCTGGAGGATTCATGACAAAACGTTTGA 771  
Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65  
Db 772 CAGCAGGGAACAAACGTAAGCGTGACGATGATCCCGAAGGTTTGTGAGCAGAA 831  
Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluileLeu 85  
Db 832 ATCCCGAAGAACTCCGCTGCTGCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTG 891  
Qy 86 -LeuCysAlaAspGlyGluProTrpLeuAla-GlyArgThrValValProValSerThrL 105  
Db 892 TGTATGTGCCGATGGCGAACC CGCGCTTGGCGGGTGGTACCGTCTTCTGTGTCAAGT 951  
Qy 105 euSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuP 125

```
Db 952 TAAGCGGCGGAGCTGGCGTTACAAAAATTGGTAAACGCCGTTAGGACGCTATCTGT 1011
Qy 125 heThrSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpG 145
Db 1012 TCACATCATCGACATTAAACCCGGACTTTATTGAGATAGGCGGTGATGCCGGCTGTGGG 1071
Qy 145 lyArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeup 165
Db 1072 GCGGACGTTCCGCGCTGCGATTAAAGCGGTAAACCGCTGTTGCTAACAGAACTGTTTTTAC 1131
Qy 165 roAlaSerProLeuTyr 170
Db 1132 CCGCGTCACCGTTGTAC 1148
```

Search completed: May 30, 2006, 03:41:56  
Job time : 1175.21 secs

**THIS PAGE BLANK (USPTO)**



Result No.	Score	Query Match	Length	DB	ID	Description
1	85.5	9.6	2907	7	US-11-293-697-1067	Sequence 1
2	80	9.0	3023	7	US-11-293-697-665	Sequence 6
3	76.5	8.6	3550	7	US-11-293-697-509	Sequence 5
4	76	8.6	1104	7	US-11-217-529-2348	Sequence 2
5	75.5	8.5	601	6	US-10-488-619-1930	Sequence 1
6	75.5	8.5	1798	7	US-11-217-529-1938	Sequence 1
7	74.5	8.4	2384	7	US-11-293-697-1066	Sequence 1
8	73	8.2	489	7	US-11-217-529-174332	Sequence 1
9	73	8.2	1406	6	US-10-953-349-34874	Sequence 3

Qy 3 ValTrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysIysGlu 22  
 ||||| :|||  
 Db 484 GTCTGGCCCTCCCATCCATCCCAAGCTG-----51

```
Qy 23 IleProAlaLeuAspProGlnLeuLeuAspTrp-----LeuLeuLeuGluAspSerMet 40
Db 511 ATCCCTACTAAGTCCACAGAGCTCTCCACCAGCAGCTCCAGAGGGGATAAATTG 570
Qy 41 ThrLysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArg-----Glu 58
Db 571 TCAAAAGCCACCAAGAGTCAGGACAGGAACCTCAGTTCTGCCAGCTCAGGGCTGTGCAG 630
Qy 59 GlyPhe-----ValGluGlnAsnGlnLeuProGluGluLeuProLeuLeuProLysGlu 76
Db 631 GGATTCAAGGTACCTGTGCGCAACCCAGCCCTCACTCATTTGAATCAGTCAGAAACCA 690
Qy 77 SerArgTyrTrpLeuArgGluLeuLeuLeuCysAlaAspGlyGluProTrpLeuAlaGly 96
Db 691 CCGCTCAGTGGATTAGAAA-----TGTGACAGGGAT-----TTCAGAGGG 732
Qy 97 ArgThrValProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGly 116
Db 733 CAATGGGTCCAGCCC-----TCACACCAGCACCACTCTGCAGGCATGGNAA 780
Qy 117 LysThrProLeuGly-Arg-----TyrLeuPheThrSerSerThrLeuThrArgAspPh 134
Db 781 CGCAGGCCAGAGGACAGGGCTTCTCAATATGGGCACACAGCAACCAAGACAGCTCCATA 840
Qy 134 eileGluLe-----GlyArgAspAlaGlyLeuTrp-----GlyAr 146
Db 841 AATAGAAGTGAATATCCAAAGCTCCCACTCCAGGGAAGCTACCACTGGGTGGATGACTC 900
Qy 146 GArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGlu 161
Db 901 CCCAGCGCTCCAAACCCCTCTCTGGCAGTCCCTTGTGCTGCCAGGAG 946

RESULT 2
US-11-293-697-665
; Sequence 665, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NOVEL full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5456
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 665
; LENGTH: 3023
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-665

Alignment Scores:
Pred. No.: 6,6 Length: 3023
Score: 80.00 Matches: 36
Percent Similarity: 39.1% Conservative: 16
Best Local Similarity: 27.1% Mismatches: 53
Query Match: 9.0% Indels: 28
DB: 7 Gaps: 5

US-10-718-311-16 (1-170) x US-11-293-697-665 (1-3023)

Qy 5 HisMetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuPro 24
Db 445 CACTTAGAAGACCACATTATAGTGAGCTGACCAACCTGAAGGTGTGATTGAATTAACA 504
Qy 25 AlaLeuAspProGlnLeuLeuAspTrpLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 44
Db 505 GGGCTCCATCTTAATAAACCAACGCCACTTGCTG-----CACCTTAGAAGACGATGG 555
Qy 45 GluGlnGln-----GlyLysThrValSer 52
|||||
```

```
Db 556 GAGCAGCAGGTGTCCGCAGCAGATGGCAAAACCTCGCGGCAAGACAGGAGGAAGTGACC 615
Qy 53 ValThrMetIleArgGluGlyPheValGluGlnAsnGluLeuProGluGluLeuPro--- 71
Db 616 CAGGCCACTCAGCCTGAGGCCATTCTCTCAGGGGACTAATCATCTACTGAAGAGAACTGGC 675
Qy 72 -----LeuLeuProLysGluSerArgTyrTrpLeuArgGluLeuLeuLeuCysAla 88
Db 676 AGGAAAGGGCAGAGGCCAAGGCCAAGAACAGTGGTCGGAAGAGTCTCTTAAACCCAGT 735
Qy 89 AspGlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeuSerThrPro 108
Db 736 GACAATGAACAGGCTTG-----CCTGTGTCTCCGGCTCTCCGCC 777
Qy 109 GluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGly 121
Db 778 -----ATGAAGAGTCTTTTCATCCACCAGTCAGGC 807

RESULT 3
US-11-293-697-509/c
; Sequence 509, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NOVEL full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 509
; LENGTH: 3550
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-509

Alignment Scores:
Pred. No.: 21,9 Length: 3550
Score: 76.50 Matches: 43
Percent Similarity: 41.1% Conservative: 15
Best Local Similarity: 30.5% Mismatches: 35
Query Match: 8.6% Indels: 49
DB: 7 Gaps: 8

US-10-718-311-16 (1-170) x US-11-293-697-509 (1-3550)

Qy 32 AspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGluGlnGlnGlyThrVal 51
Db 397 GACACACCTTAAGTGAAGAGTGCCTCATACAGAGTCCACAGAGGGCAAT--GCG 340
Qy 52 SerValThrMetIleArgGluGlyPheValGluGln-----Asn 64
|||||
Db 339 AGTGTG-----GAGGGAAGCATCTCCAGGATTCAGGGCGCCCGCTGACGAC 292
Qy 65 GluLeuProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLe 84
Db 291 TCCTGCCCGGACAAAGGTCCTCCAGGATGAGAGCGATGAGCGCCCTGACCCATC 232
Qy 85 -----LeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValPro--- 101
Db 231 GTCAGGTTGACCTGC-----TGG-----CCAACCTCAGTTTCTCCAAGG 193
Qy 102 ValSerThrLeuSerGlyProGluLeuAla----- 111
Db 192 GTCACATGCTCTTTGGTCTCTGAGTTACCCCTCATCGGAGACGAGAGGCATGGAGGAAA 133
Qy 112 -----LeuGlnLysLeuGlyLysThrProLeu 120
Db 132 AGGGGGTACAGACCCCGAGGAGGAGTTGGCTTCAGACAGAAAGACAAAGAACCCGGGTG 73
Qy 121 GlyArgTyrLeuPhe-----ThrSerSerThrLeuThrArgAspPhe 134
```





```

; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1629
; LENGTH: 2019
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-1629

Alignment Scores:
Pred. No.:      25.4      Length:      2019
Score:          73.00     Matches:      51
Percent Similarity: 33.2%  Conservative: 20
Best Local Similarity: 23.8% Mismatches:   49
Query Match:     8.2%     Indels:       94
DB:              7       Gaps:          9

US-10-718-311-16 (1-170) x US-11-293-697-1629 (1-2019)

QY      1 MetGlnValTrpHisMetSerHisPro-----AlaLeuThrGlnLeu 14
Db      1968 TTAAAGATGTGGAAGATGGGTACAAATTAACCATGAGAGTTGTGCAGGGAACACCCGT 1909

QY      15 ArgAlaLeuArgTyxCysLysGluLeuProAla-----LeuAspProGlnLeu 31
Db      1908 AGGGCTGTGTCACCTTCAGATATGCTGCTCCCAAAATTCAGACCCACAGATG--- 1852

QY      32 AspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGluGlnGlnGlyLysThrVal 51
Db      1851 -----ATAGAAGGGTGAGTCAAGCAAGGAGAGGCCTCTCTGTAGAGGCTG 1792

QY      52 SerValThrMetIleArgGluGlyPheValGluGlnAsnGluLeuPro----- 67
Db      1839 -----ATAGAAGGGTGAGTCAAGCAAGGAGAGGCCTCTCTGTAGAGGCTG 1792

QY      67 ----- 67
Db      1791 AGGTCCCTCTGTTTCAAGGATGGATGTACACCTTGACCTTCGGGGTCTGCAGTGGC 1732

QY      68 GluGluLeuProLeuLeuProLysGluSerArgTyTrp---LeuArgGluLeuLeuLeu 86
Db      1731 CAGAGGCTGCTGTGCGCCCTTCCTCTCCCTTCTTGGGCACTGTGGAGGCTCTCTTTG 1672

QY      87 CysAlaAspGly-GluProTrpLeuAlaGlyArgThrValProValSerThrLeuSe 106
Db      1671 CTGTGTGGAGGTGTGTCC----- 1653

QY      106 rGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrPro-LeuGlyArgTyTrpLeuPhe 126
Db      1652 -----CTCAGGCCGTGGGTGAGTCTCTCTAGGAGGCTCTCTTTG 1609

QY      126 hrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGlyA 146
Db      1608 CC-----CTTCGTGTCTCGAAGGGGCTTGTCTGGAGGCA 1573

QY      146 rArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeu---- 164
Db      1572 AAGCGTCTCCACTCTGTCTCTCAGACTCAGCTGTGTGGCCTTGATTTCTTTTGGCGG 1513

QY      165 -----ProAlaSerPro 168
Db      1512 ACTTGGCCCTTGTGGTGCCACCGGTCCAGGATCCGCC 1475

RESULT 11
US-10-524-433-1
; Sequence 1, Application US/10524433
; Publication No. US20060099676A1
```

```

; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1629
; LENGTH: 2019
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-1629

Alignment Scores:
Pred. No.:      25.4      Length:      2019
Score:          73.00     Matches:      51
Percent Similarity: 33.2%  Conservative: 20
Best Local Similarity: 23.8% Mismatches:   49
Query Match:     8.2%     Indels:       94
DB:              7       Gaps:          9

US-10-718-311-16 (1-170) x US-11-293-697-1629 (1-2019)

QY      1 MetGlnValTrpHisMetSerHisPro-----AlaLeuThrGlnLeu 14
Db      1968 TTAAAGATGTGGAAGATGGGTACAAATTAACCATGAGAGTTGTGCAGGGAACACCCGT 1909

QY      15 ArgAlaLeuArgTyxCysLysGluLeuProAla-----LeuAspProGlnLeu 31
Db      1908 AGGGCTGTGTCACCTTCAGATATGCTGCTCCCAAAATTCAGACCCACAGATG--- 1852

QY      32 AspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGluGlnGlnGlyLysThrVal 51
Db      1851 -----ATAGAAGGGTGAGTCAAGCAAGGAGAGGCCTCTCTGTAGAGGCTG 1792

QY      52 SerValThrMetIleArgGluGlyPheValGluGlnAsnGluLeuPro----- 67
Db      1839 -----ATAGAAGGGTGAGTCAAGCAAGGAGAGGCCTCTCTGTAGAGGCTG 1792

QY      67 ----- 67
Db      1791 AGGTCCCTCTGTTTCAAGGATGGATGTACACCTTGACCTTCGGGGTCTGCAGTGGC 1732

QY      68 GluGluLeuProLeuLeuProLysGluSerArgTyTrp---LeuArgGluLeuLeuLeu 86
Db      1731 CAGAGGCTGCTGTGCGCCCTTCCTCTCCCTTCTTGGGCACTGTGGAGGCTCTCTTTG 1672

QY      87 CysAlaAspGly-GluProTrpLeuAlaGlyArgThrValProValSerThrLeuSe 106
Db      1671 CTGTGTGGAGGTGTGTCC----- 1653

QY      106 rGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrPro-LeuGlyArgTyTrpLeuPhe 126
Db      1652 -----CTCAGGCCGTGGGTGAGTCTCTCTAGGAGGCTCTCTTTG 1609

QY      126 hrSerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGlyA 146
Db      1608 CC-----CTTCGTGTCTCGAAGGGGCTTGTCTGGAGGCA 1573

QY      146 rArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeu---- 164
Db      1572 AAGCGTCTCCACTCTGTCTCTCAGACTCAGCTGTGTGGCCTTGATTTCTTTTGGCGG 1513

QY      165 -----ProAlaSerPro 168
Db      1512 ACTTGGCCCTTGTGGTGCCACCGGTCCAGGATCCGCC 1475

RESULT 11
US-10-524-433-1
; Sequence 1, Application US/10524433
; Publication No. US20060099676A1
```

```

; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 34874
; LENGTH: 1406
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
US-10-953-349-34874

Alignment Scores:
Pred. No.:      15      Length:      1406
Score:          73.00     Matches:      29
Percent Similarity: 42.0%  Conservative: 21
Best Local Similarity: 24.4% Mismatches:   35
Query Match:     8.2%     Indels:       34
DB:              6       Gaps:          4

US-10-718-311-16 (1-170) x US-10-953-349-34874 (1-1406)

QY      56 IleArgGluGlyPheValGluGlnAsnGluIlePro-----GluGlu 69
Db      906 CTGAGTCAGCAGCGTAGCTGTTCACAGTTGCCGATCTCGCGGGGATGACGAGGAG 847

QY      70 LeuProLeuLeuProLysGluSerArgTyTrpLeuArgGluIleLeuLeuCysAlaAsp 89
Db      846 ATGCCGCTGTGAGGAGG----- 829

QY      90 GlyGluProTrpLeuAlaGlyArgThrValProValSerThrLeuSerGlyProGlu 109
Db      828 -----ATGAGCTCGTGTGACGTGCGCGGCC 805

QY      110 LeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyTrpLeuPheThrSerThr 129
Db      804 ATGTGCCCATGCTGCTCGGAGCGAGCGCTCACCGCGGAGGTTG---GCCAGCAGC 748

QY      130 LeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGlyArgSerArg 149
Db      747 ATCACC---GACGCCGTGAGTTGGTGAAGCTGGCGGCGAGCGTGAACCTGGAAGTGGTTG 691

QY      150 LeuArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuProAlaSerPro 168
Db      690 TTGTTGATGAAAGCGCGTCGATCTTCTTGTCAAGATGCGCGCGGACCTCGCG 634

RESULT 10
US-11-293-697-1629/c
; Sequence 1629, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
```

```

; GENERAL INFORMATION:
; APPLICANT: Functional Genetics, Inc.
; TITLE OF INVENTION: MAMMALIAN GENES INVOLVED IN RAPAMYCIN
; FILE REFERENCE: 10784-023-228
; CURRENT APPLICATION NUMBER: US/10/524,433
; CURRENT FILING DATE: 2005-02-15
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 966
; TYPE: DNA
; ORGANISM: murine
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 548, 564, 567, 583, 603, 610, 636, 647, 648, 651, 654, 656,
; LOCATION: 661, 667, 669, 675, 691, 710, 713, 719, 732, 746, 749, 754,
; LOCATION: 761, 771, 777, 781, 792, 795, 802, 808, 812, 813, 824, 833,
; LOCATION: 841, 842, 856, 861, 866, 871, 873, 874, 875, 882, 887
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 897, 902, 905, 910, 915, 923, 927, 928, 935, 938, 948, 949,
; LOCATION: 965, 966
; OTHER INFORMATION: n = A,T,C or G
US-10-524-433-1

Alignment Scores:
Pred. No.: 13.2      Length: 966
Score: 71.50      Matches: 33
Percent Similarity: 40.8%      Conservative: 18
Best Local Similarity: 26.4%      Mismatches: 56
Query Match: 8.1%      Indels: 19
DB: 6      Gaps: 4

US-10-718-311-16 (1-170) x US-10-524-433-1 (1-966)
Qy 44 PheGluGlnGlnGlyThrValSerValThrMetIleArgGluGlyPheValGluGln 63
Db 385 TTTAAACACCAAGGAAAG---TACACAGCCTCCCGCCTCAGCGGAGCCAGTGGAAAGT 441
Qy 64 AsnGluIleProGluLeuProLeuLeuProLysGlySerArgTyrTrpLeuArgGlu 83
Db 442 AGCCACCTTCCTGCAATCTGCCACACTGGCAGAGAGAGCCTGCATGCCCGGTGG 501
Qy 84 IleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValProValSer 103
Db 502 GCTTTGG-TGC-----TGGCTCTGTACTCCCACTTGTCTGTGTGGCC 544
Qy 104 ThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyr 123
Db 545 TCANGAAAGGT---CATCTCTNAANGAAACAAAGCCACCTNCTCCCTCCCTGGCATG 601
Qy 124 LeuPheThrSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeu 143
Db 602 CNTTTTCNGCCACAGCCCGTGTGTCCCAACANAATTTGGNCANGANGNAGGN 661
Qy 144 TrpGlyArgSerArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPhe 163
Db 662 TGGGANAAACCCNAACCTA-----TTGTGG 688
Qy 164 LeuProAlaSerPro 168
Db 689 CTNCCCTGTGCCCC 703

RESULT 12
US-11-293-697-200/c
; Sequence 200, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106

```

```

; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 200
; LENGTH: 2490
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-200

Alignment Scores:
Pred. No.: 52      Length: 2490
Score: 71.50      Matches: 34
Percent Similarity: 45.4%      Conservative: 15
Best Local Similarity: 31.5%      Mismatches: 30
Query Match: 8.1%      Indels: 29
DB: 7      Gaps: 9

US-10-718-311-16 (1-170) x US-11-293-697-200 (1-2490)
Qy 27 AspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGluGln 46
Db 755 GACCTGGATAGGTAATTGGCTCTCTCTGCC-----TCAGAATCCCTTCAGCAG 705
Qy 47 GlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGluIle 66
Db 704 CAGGACCTCACATG-----GGCTTTGTG-----GTC 678
Qy 67 ProGluGlu-----LeuProLeu---LeuProLysGlySerArgTyrTrpLeuArgGlu 83
Db 677 CCTGAGGAACCTTCCTGTTCTGGTTTGCATCCAGATGGCTGTGGCGGTGGAG 618
Qy 84 IleLeuLeuCysAlaAspGlyGluPro-----TrpLeuAlaGlyArgThrVal 99
Db 617 CTGGGGATCCACACATTCTTCAGCCACCTGGCAATTAATGGAGGCTGAGAAGTCCCTG 558
Qy 100 -----ValProValSerThrLeuSerGlyProGluLeuAlaLeuGln 113
Db 557 AGAGACTATTTCGAAGAGGTGCCAGTAGCGTGGCTCAGT---CCAAGTCTCAAGCCTCA 501
Qy 114 LysLeuGlyLysThrProLeuGly 121
Db 500 GAACACAGGG---ACGCCACTGGGA 480

RESULT 13
US-10-473-691B-1/c
; Sequence 1, Application US/10473691B
; Publication No. US20060099202A1
; GENERAL INFORMATION:
; APPLICANT: Nicolette, Charles A.
; APPLICANT: Solitis, Daniel A.
; TITLE OF INVENTION: IMMUNOGLOBULIN CONSTRUCT CONTAINING TUMOR SPECIFIC p53BP2 SEQUEN
; TITLE OF INVENTION: FOR ELICITING AN ANTI-TUMOR RESPONSE
; FILE REFERENCE: 02755/100K313-US1
; CURRENT APPLICATION NUMBER: US/10/473,691B
; CURRENT FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: PCT/US02/10224
; PRIOR FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: US 60/280,733
; PRIOR FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 4534
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (757) ..(3774)
; OTHER INFORMATION:
; PUBLICATION INFORMATION:

```

; AUTHORS: Naumovski, L. and Cleary, M.L.  
 ; TITLE: The P53-binding protein 53BP2 also interacts with Bcl2 and impedes  
 ; TITLE: cell cycle progression at G2/M  
 ; JOURNAL: Mol. Cel. Biol.  
 ; VOLUME: 16  
 ; ISSUE: 7  
 ; PAGES: 3884-3892  
 ; DATE: 1996-07  
 ; DATABASE ACCESSION NUMBER: Genbank / US8334  
 ; DATABASE ENTRY DATE: 1996-07-02  
 US-10-473-691B-1

Alignment Scores:  
 Pred. No.: 124 Length: 4534  
 Score: 71.50 Matches: 71.28  
 Percent Similarity: 39.2% Conservative: 12  
 Best Local Similarity: 27.5% Mismatches: 33  
 Query Match: 8.1% Indels: 29  
 DB: 6 Gaps: 5

US-10-718-311-16 (1-170) x US-10-473-691B-1 (1-4534)  
 QY 68 GluGluLeuProLeuLeuProLysGluSerArgTyr----- 79  
 Db 1670 GAAGCAGAGCCCTGGCTGGGAAAGATCTGCATTGGAAGGACTCCAATCAGGGCCAACT 1611  
 QY 80 --TTPLeuArgGluIleLeuLeuCysAlaAspGlyGluProTPrLeuAlaGlyArgThr 98  
 Db 1610 GGTGATTTAGACCTTAGTTGTGTGAGCAGCCCGCCAGATCTCATGTGGGCAGTGTC 1551  
 QY 99 ValValProValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThr 118  
 Db 1550 TGTATTTTCATC-----GGCCCTCTGAAGCC-----TGAATGACCAAGGAA 1509  
 QY 119 ProLeuGlyArgTyrLeuPheThrSerThr-----LeuThr 131  
 Db 1508 CCATCCGCGAGGCTGGCTTACCAGCAATTCAGGCCTTGGGGCATCGAGGCATAGTA 1449  
 QY 132 ArgAspPheIle-----GluIleGlyArgAspAlaGlyLeuTPr 144  
 Db 1448 GATGACTGGATATAGGACCTACTGCAGCCACACGCGTTGGGGCTGACCGCGCTTGCTGG 1389  
 QY 145 GlyArg 146  
 Db 1388 GGAAGA 1383

RESULT 14  
 US-11-293-697-336  
 ; Sequence 336, Application US/11293697  
 ; Publication No. US20060105376A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HELIX RESEARCH INSTITUTE  
 ; TITLE OF INVENTION: Novel full length cDNA  
 ; FILE REFERENCE: HI-A0106  
 ; CURRENT APPLICATION NUMBER: US/11/293,697  
 ; CURRENT FILING DATE: 2005-12-05  
 ; PRIOR APPLICATION NUMBER: US/10/108,260  
 ; PRIOR FILING DATE: 2002-03-28  
 ; NUMBER OF SEQ ID NOS: 5458  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 336  
 ; LENGTH: 3105  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-11-293-697-336

Alignment Scores:  
 Pred. No.: 82.2 Length: 3105  
 Score: 71.00 Matches: 32  
 Percent Similarity: 43.6% Conservative: 9  
 Best Local Similarity: 34.0% Mismatches: 35  
 Query Match: 8.0% Indels: 18  
 DB: 7 Gaps: 4

US-10-718-311-16 (1-170) x US-11-293-697-336 (1-3105)  
 QY 43 ArgPheGluGln-----GlnGlyLysThrValSerValThrMetIleArgGlu 58  
 Db 1933 CGTTTTTGACCATCAATGCTGTGGATGAACATGCAATGCTCAATGTCAGAAAGAAAGAG 1992  
 QY 59 GlyPheValGluGlnAsnGluIleProGluGluLeuProLeuLeuProLysGluSerArg 78  
 Db 1993 AATTTTGGCCATCAAAATCCAGCTCAAGTGAACCT-----AAAGAGCTTATA 2040  
 QY 79 TyrTrpLeuArgGluIleLeuLeuCysAlaAspGlyGluProTPrLeuAlaGlyArg--- 97  
 Db 2041 TATTTTCATGAAGCTGATCCTTTTGTGTGTGTGCTCC-----TGGTGTTAGGAGAGAAAA 2094  
 QY 98 -----ThrValValProValSerThrLeuSerGlyProGluLeu 110  
 Db 2095 AAGCTCTATGAAGAATATAGGAAGTTCTCTCTTTTCACACCTTATTTTCATTGACTGCTG 2154  
 QY 111 AlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeu 124  
 Db 2155 GCTGCTTTAAAAAAGGATCAATGTACCGTTGCTTT 2196

## RESULT 15

US-10-488-619-48/c  
 ; Sequence 48, Application US/10488619  
 ; Publication No. US20060099578A1  
 ; GENERAL INFORMATION:

; APPLICANT: Greenlee, Winner and Sullivan, P.C.  
 ; TITLE OF INVENTION: Human Mitochondrial DNA Polymorphisms, Haplogroups, Associations  
 ; TITLE OF INVENTION: Physiological Conditions, And Genotyping Arrays  
 ; FILE REFERENCE: 98-01 WO  
 ; CURRENT APPLICATION NUMBER: US/10/488,619  
 ; CURRENT FILING DATE: 2004-03-01  
 ; NUMBER OF SEQ ID NOS: 3040  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 48  
 ; LENGTH: 674  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)..(674)  
 ; OTHER INFORMATION: n is g, c, a or t  
 US-10-488-619-48

## Alignment Scores:

Pred. No.: 10.3 Length: 674  
 Score: 70.50 Matches: 53  
 Percent Similarity: 37.4% Conservative: 14  
 Best Local Similarity: 29.6% Mismatches: 59  
 Query Match: 7.9% Indels: 53  
 DB: 6 Gaps: 10

US-10-718-311-16 (1-170) x US-10-488-619-48 (1-674)

QY 2 GlnValTrpHisMetSerHisPro-AlaLeuThrGlnLeuArgAlaLeuArgTyrCysIly 21  
 Db 498 CAGCGCTGGCTACCGCAGCCATCCAGCCCTGCTCAGCAGCCGAGAGGCTGGATCCGAGGAGGT 447  
 QY 21 sGluIleProAlaLeuAspTrpLeuLeuAspTrpLeuLeuLeuLeuLeuLeuLeuLeu 41  
 Db 446 -----CCAGTCTCTTTCTCAGACTGGGAGAGAGCTGGATCCGAGGAGGT 403  
 QY 41 rLysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheVa 61  
 Db 402 GGGCCGGGGCCAGGGCACGGGAGAGCC-----AGGAGAGAGCTGCT 361  
 QY 61 lGluGlnAsnGluIle-----ProGluGluLeuProIle 72  
 Db 360 GGATCTCAGGAGATGCTGCGCTCTCTGGGCCACTGAGCCAGCCCGCCGCCCCCCCC 301  
 QY 72 uLeuProLys-----GluSerArgTyrTrpLeuArgGlu-IleLeuLeuCys 87

```

Db      300 AGCAGGGAAGGATGAGTGTGGGAGGGAAGGGCTGGTCCGTCTGATGGGACTTTGC 241
Qy      88 -----AlaAsp---GlyGluProTyrLeuAlaGlyArgThrValValProValSerThr 104
Db      240 ACCTCTGCTGATCCGGCCGGCCCTGGCTTGGAGGCTTGGCTGCTCTTCCAGCGTCT--- 184
Qy      105 LeuSerGlyProGluLeuAlaLeu-GlnLysLeuGlyLysThrProLeuGlyArgTyrLe 124
Db      183 -----CTCCTCCCTCTCTGGGAAGGTGCGCCCTTGGCGCGCAAGGT 145
Qy      124 u-----PheThrSerSerThrLeuThrArgAspPheIleGluIleGly----- 138
Db      144 TTTAGCTTTCAGCAACTGAGTAACCTTAGGCACAGGTGGAGGTGTGGCCCGATCTAAC 85
Qy      139 -----ArgAspAlaGlyLeuTyrGlyArgArgSerArgLeu 150
Db      84 CCTTACCCATCTCTACTGTGACTGTGG---AGGGTCACCAAGTTG 39
```

Search completed: May 30, 2006, 05:54:13  
Job time : 42.0299 secs



GenCore version 5.1.8  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 30, 2006, 01:43:54 ; Search time 4389.04 Seconds  
(without alignments)  
3248.869 Million cell updates/sec

Title: US-10-718-311-16

Perfect score: 887

Sequence: 1 MQVWMSHPALQRLALRYC.....RLSGKPLLTLELFPASPLY 170

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 48236798 seqs, 27959665780 residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=/abs/ABSSWEB spool/US10718311/runat\_26052006\_164919\_13483/app\_query.fasta\_1  
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -IOPCU=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs02h  
-USER=US10718311@CGN 1.1.7986 @runat\_26052006\_164919\_13483 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database :

EST:.\*  
1: gb\_est1:.\*  
2: gb\_est3:.\*  
3: gb\_est4:.\*  
4: gb\_est5:.\*  
5: gb\_est6:.\*  
6: gb\_hic:.\*  
7: gb\_est2:.\*  
8: gb\_est7:.\*  
9: gb\_est8:.\*  
10: gb\_est9:.\*  
11: gb\_gss1:.\*  
12: gb\_gss2:.\*  
13: gb\_gss3:.\*  
14: gb\_gss4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	618	69.7	832	13	CL666350 PRI0152b_
2	508	57.3	828	13	CL687382 PRI0146c_
3	262	29.5	536	14	AG192190 Pan trogl
4	236	26.6	533	14	AG192116 Pan trogl

5	209	23.6	515	12	CC048367
6	183	20.6	563	13	CL650764
7	171	19.3	838	13	CL688660
8	160	18.0	440	14	AG192305
9	150	16.9	1110	11	BZ551713
c 10	138	15.6	862	11	BZ564576
c 11	118	13.3	739	14	AJ863670
c 12	105	11.8	1141	12	BZ578851
c 13	95	10.7	753	13	CL696699
c 14	93	10.5	581	2	BG799182
c 15	92.5	10.4	4527	6	CR861280
c 16	92	10.4	1065	13	CW917940
c 17	91	10.3	666	2	BJ932163
c 18	91	10.3	1033	14	DU782699
c 19	90.5	10.2	668	10	DY396666
c 20	90	10.1	826	8	CO384679
c 21	90	10.1	1888	6	AY850332
c 22	89	10.0	451	3	BM870249
c 23	89	10.0	528	12	CE278012
c 24	89	10.0	770	7	BE618275
c 25	88.5	10.0	674	8	CX288046
c 26	88.5	10.0	805	8	CV120546
c 27	88.5	10.0	920	8	CO161709
c 28	88	9.9	551	9	DA934706
c 29	88	9.9	785	3	BQ111035
c 30	88	9.9	895	3	BQ642631
c 31	87.5	9.9	797	14	DU809059
c 32	87.5	9.9	931	3	BQ550781
c 33	87.5	9.9	1114	5	CK209417
c 34	87.5	9.9	2621	6	AK004672
c 35	87.5	9.9	2929	6	AK158339
c 36	87.5	9.9	2933	6	AK157854
c 37	87	9.8	705	2	BJ578705
c 38	87	9.8	717	2	BJ566157
c 39	87	9.8	746	13	CL393085
c 40	87	9.8	959	3	BU513771
c 41	87	9.8	988	5	CF513983
c 42	86.5	9.8	422	7	BF543156
c 43	86.5	9.8	615	2	BM265186
c 44	86.5	9.8	750	2	BJ974078
c 45	86.5	9.8	788	12	BZ578446

#### ALIGNMENTS

RESULT 1  
CL666350/c

LOCUS  
DEFINITION

CL666350  
PRI0152b\_A03 - PRI0152b.B21 (832) Mixed stage foamid library of P.  
pacificus var. California Pristionchus pacificus genomic, genomic  
survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

COMMENT

CL666350.1 GI:50159025  
GSS.  
Pristionchus pacificus  
Pristionchus pacificus  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.  
1 (bases 1 to 832)  
Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.  
AppaDB: an AcedB database for the nematode satellite organism  
Pristionchus pacificus  
Nucleic Acids Res. 32 (1), D421-D422 (2004)  
14881447  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498  
Email: raif.sommer@tuebingen.mpg.de  
This library was generated at Caltech, Pasadena, USA and end  
sequenced at Vancouver, Canada.

Seq primer: T7

Class: fosmid ends.

Location/Qualifiers

## FEATURES

source

1. 832  
 /organism="Pristionchus pacificus"  
 /mol\_type="genomic DNA"  
 /strain="California"  
 /db\_xref="taxon:54126"  
 /clone\_lib="Mixed stage fosmid library of P. pacificus var. California"  
 /note="Vector: pEpifos-5 Fosmid vector"

## ORIGIN

Alignment Scores:  
 Pred. No.: 1.24e-57 Length: 832  
 Score: 618.00 Matches: 120  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 69.7% Indels: 0  
 DB: 13 Gaps: 0

US-10-718-311-16 (1-170) x CL666350 (1-832)

Qy 51 ValSerValThrMetIleArgGluGlyPheValGluGlnAsnGluIleProGluGluLeu 70  
 |||||  
 Db 832 GTAACGGTGCAGATGATCCGAGAGGTTTGTGAGCAGATGAAATCCCGAAGACTG 773  
 |||||  
 Qy 71 ProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeuLeuLysAlaAspGly 90  
 |||||  
 Db 772 CGCTGTCTGCCGAAAGAGTCTCGTTACTGTTACGTGAAATTTGTTATGTCCGATGGT 713  
 |||||  
 Qy 91 GluProTrpLeuAlaGlyArgThrValProValSerThrLeuSerGlyProGluLeu 110  
 |||||  
 Db 712 GAACCGTGGCTTGGCGGTGTCACCGTCTCTGTGTCAACGTTTAAGCGGCGGAGCTG 653  
 |||||  
 Qy 111 AlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPheThrSerSerThrLeu 130  
 |||||  
 Db 652 GCGTTACAAAATTTGGTAAACGCCGTTAGACGCTATCTGTTACATCATCTGCACATTA 593  
 |||||  
 Qy 131 ThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGlyArgSerArgLeu 150  
 |||||  
 Db 592 ACCCGGACTTTATTGAGATAGCGCGTATGCCGGCTGTGGGGCGAGCTTCCGCGCTG 533  
 |||||  
 Qy 151 ArgLeuSerGlyLysProLeuLeuLeuThrGluLeuPheLeuProAlaSerProLeuTyr 170  
 |||||  
 Db 532 CGATTACCGGTAAACCGCTGTGTCTTAACAGAACTGTTTTTACCAGCGCTCACCGTTGTAC 473  
 |||||

## RESULT 2

CL687382

LOCUS

DEFINITION CL687382 828 bp DNA linear GSS 09-JUL-2004  
 PRI0146c\_H03\_2 - PRI0146c-BR (828) Mixed stage fosmid library of P.  
 pacificus var. California Pristionchus pacificus genomic, genomic  
 survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Pristionchus pacificus

Pristionchus pacificus

Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;

Neodiplogasteridae; Pristionchus.

1 (bases 1 to 828)

Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.

AppaB: an AcedB database for the nematode satellite organism

Pristionchus pacificus

Nucleic Acids Res. 32 (1), D421-D422 (2004)

14681447

Contact: Sommer RJ

Evolutionary Biology

Max-Planck-Institute for Developmental Biology

Spemannstr. 37-39, Tuebingen D-72076, Germany

Tel: 00497071601371

Fax: 00497071601498

Email: ralf.sommer@tuebingen.mpg.de

This library was generated at Caltech, Pasadena, USA and end  
 sequenced at Vancouver, Canada.

Seq primer: T7

Class: fosmid ends.

Location/Qualifiers

## FEATURES

source

1. 828  
 /organism="Pristionchus pacificus"  
 /mol\_type="genomic DNA"  
 /strain="California"  
 /db\_xref="taxon:54126"  
 /clone\_lib="Mixed stage fosmid library of P. pacificus var. California"  
 /note="Vector: pEpifos-5 Fosmid vector"

## ORIGIN

Alignment Scores:  
 Pred. No.: 1.93e-45 Length: 828  
 Score: 508.00 Matches: 98  
 Percent Similarity: 99.0% Conservative: 0  
 Best Local Similarity: 99.0% Mismatches: 1  
 Query Match: 57.3% Indels: 0  
 DB: 13 Gaps: 0

US-10-718-311-16 (1-170) x CL687382 (1-828)

Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25  
 |||||  
 Db 532 ATGTGCACACCCCGGTTAAACGCAACTGCGTGCCTATTTTAAAGAGATCCCTGCC 591  
 |||||  
 Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGlu 45  
 |||||  
 Db 592 CTGATCCGCAACACTGCTCGACTGGCTGTGCTGAGGATTCATGACAAAACGTTTGA 651  
 |||||  
 Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGlnGlnAsnGlu 65  
 |||||  
 Db 652 CAGCAGCGAAAACCGTAAGCGTACCGATCCGCGAAGGGTTTGTCCGACAGAAATGA 711  
 |||||  
 Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85  
 |||||  
 Db 712 ATCCCCGAAGAACTGCCGCTGCTGCCGAAAGAGTCTCGTTACTGGTTACGTGAAAATTTG 771  
 |||||  
 Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThr 104  
 |||||  
 Db 772 TTATGTCGATGTGTAACCGTGTGCTGCTGCTACCGTCTGCTCTCTGTCGTAACG 828  
 |||||

AG192190 536 bp DNA linear GSS 06-MAR-2004

LOCUS

DEFINITION

AG192190 sequence.  
 AG192190  
 AG192190.1 GI:45224366  
 GSS.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Pan troglodytes (chimpanzee)

Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Pan.

1 (bases 1 to 536)

Park,H., Kim,Y., Kim,S., Han,Y., Woo,T., Park,K., Eun,C.J.,

Hoon,S.T., Chu,M., Kim,H., Joo,S., Kim,C., Song,W. and Yoo,H.

BAC end sequences of Library RP-43

Unpublished

2 (bases 1 to 536)

Park,H., Kim,Y., Kim,S., Han,Y., Woo,T., Park,K., Eun,C.J.,

Hoon,S.T., Chu,M., Kim,H., Joo,S., Kim,C., Song,W. and Yoo,H.

Direct Submission

Submitted (07-JAN-2002) Hong-Seog Park, Korea Research Institute of

Bioscience and Biotechnology (KRIIB), Genome Research Center (GRC);

52, Oun-dong, Yusong-gu, Daejeon 305-333, Korea

(E-mail: redstone@mail.kribb.re.kr, URL:http://phs.grc.kribb.re.kr/.

Tel:82-42-866-7181, Fax:82-42-860-4409)

Clones are derived from the chimpanzee BAC library RP-43 This BAC

COMMENT



/note="Vector: TOPO-PCR4; DNA flanking Mu transposon insertions in Mu inactive lines were extracted from the UniformMu maize population by the thermo asymmetric interlaced PCR (TAIL) protocol using primers specific for the Mu terminal inverted repeat and a set of 16 arbitrary primers. Amplicons were size enriched using Sepharose 400 spin columns and cloned into the TOPO PCR4 vector."

## ORIGIN

Alignment Scores:  
 Pred. No.: 1.36e-12 Length: 515  
 Score: 209.00 Matches: 40  
 Percent Similarity: 97.6% Conservative: 0  
 Best Local Similarity: 97.6% Mismatches: 1  
 Query Match: 23.6% Indels: 0  
 DB: 12 Gaps: 0

US-10-718-311-16 (1-170) x CC048367 (1-515)

Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 25  
 |||||  
 Db 393 ATGTCACACCCCGCTTAACGCACTGCGTGGCTGCGCTATTGTACAGAGATCCCTGCC 452  
 |||||  
 Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45  
 |||||  
 Db 453 CTGGATCCGCAACTGCTGACTGGCTGTGTGGAGGATTCATGACAAAACGTTTGA 512  
 |||||  
 Qy 46 Gln 46  
 |||||  
 Db 513 CAA 515

RESULT 6  
 CL650764  
 LOCUS  
 DEFINITION PR10110b.P06 - PRI0110b.B21 (563) Mixed stage fosmid library of P. pacificus var. California Pristionchus pacificus genomic, genomic survey sequence.  
 CL650764 563 bp DNA linear GSS 09-JUL-2004  
 PR10110b.P06 - PRI0110b.B21 (563) Mixed stage fosmid library of P. pacificus var. California Pristionchus pacificus genomic, genomic survey sequence.

ACCESSION CL650764  
 VERSION CL650764.1 GI:50129323  
 KEYWORDS GSS.

SOURCE  
 ORGANISM Pristionchus pacificus  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.

REFERENCE 1 (bases 1 to 563)  
 Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.  
 AppaDB: an AcedB database for the nematode satellite organism Pristionchus pacificus  
 Nucleic Acids Res. 32 (1), D421-D422 (2004)

JOURNAL  
 PUBMED 14681447

COMMENT  
 Contact: Sommer RJ  
 Evolutionary Biology  
 Max-Planck-Institute for Developmental Biology  
 Spemannstr. 37-39, Tuebingen D-72076, Germany  
 Tel: 00497071601371

Fax: 00497071601498  
 Email: raf.f.sommer@uebingen.mpg.de  
 This library was generated at Caltech, Pasadena, USA and end sequenced at Vancouver, Canada.  
 Seq primer: T7  
 Class: fosmid ends.

Location/Qualifiers  
 1..563

FEATURES  
 source

/organism="Pristionchus pacificus"  
 /mol\_type="genomic DNA"  
 /strain="California"  
 /db\_xref="taxon:54126"  
 /clone\_lib="Mixed stage fosmid library of P. pacificus var. California"  
 /note="Vector: pBpifos-5 Fosmid vector"

## ORIGIN

Alignment Scores:

Pred. No.: 1.18e-09 Length: 563  
 Score: 183.00 Matches: 38  
 Percent Similarity: 97.4% Conservative: 0  
 Best Local Similarity: 97.4% Mismatches: 0  
 Query Match: 20.6% Indels: 1  
 DB: 13 Gaps: 0

US-10-718-311-16 (1-170) x CL650764 (1-563)

Qy 133 AspPheileGluileGlyArgAspAlaGlyLeu-TrpGlyArgSerArgLeuArgLe 152  
 |||||  
 Db 2 GACTTTATTGAGATAGCGCGTGATGCCGGCTCGTGGGGCGACGTTCCCGCTCGGATT 61  
 |||||  
 Qy 152 uSerGlyLysProLeuLeuLeuThrGluLeuPheLeuProAlaSerProLeuTyr 170  
 |||||  
 Db 62 AAGCGGTAACCGCTGTTGCTTAACAGAACTGTTTTTACCGCGCTCACCGTTGTAC 116  
 |||||

## RESULT 7

CL688660  
 LOCUS  
 DEFINITION PR1014a.D06.2 - PRI014a.BR (838) Mixed stage fosmid library of P. pacificus var. California Pristionchus pacificus genomic, genomic survey sequence.  
 CL688660 838 bp DNA linear GSS 09-JUL-2004  
 PR1014a.D06.2 - PRI014a.BR (838) Mixed stage fosmid library of P. pacificus var. California Pristionchus pacificus genomic, genomic survey sequence.

ACCESSION CL688660  
 VERSION CL688660.1 GI:50197738  
 KEYWORDS GSS.

SOURCE  
 ORGANISM Pristionchus pacificus  
 Pristionchus pacificus  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.

REFERENCE 1 (bases 1 to 838)  
 Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.  
 AppaDB: an AcedB database for the nematode satellite organism Pristionchus pacificus  
 Nucleic Acids Res. 32 (1), D421-D422 (2004)

JOURNAL  
 PUBMED 14681447

COMMENT  
 Contact: Sommer RJ  
 Evolutionary Biology  
 Max-Planck-Institute for Developmental Biology  
 Spemannstr. 37-39, Tuebingen D-72076, Germany  
 Tel: 00497071601371

Fax: 00497071601498  
 Email: raf.f.sommer@uebingen.mpg.de  
 This library was generated at Caltech, Pasadena, USA and end sequenced at Vancouver, Canada.  
 Seq primer: T7  
 Class: fosmid ends.

Location/Qualifiers  
 1..838

FEATURES  
 source

/organism="Pristionchus pacificus"  
 /mol\_type="genomic DNA"  
 /strain="California"  
 /db\_xref="taxon:54126"  
 /clone\_lib="Mixed stage fosmid library of P. pacificus var. California"  
 /note="Vector: pBpifos-5 Fosmid vector"

## ORIGIN

Alignment Scores:

Pred. No.: 4.46e-08 Length: 838  
 Score: 171.00 Matches: 36  
 Percent Similarity: 94.7% Conservative: 0  
 Best Local Similarity: 94.7% Mismatches: 1  
 Query Match: 19.3% Indels: 1  
 DB: 13 Gaps: 0

US-10-718-311-16 (1-170) x CL688660 (1-838)

Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyr-CysLysGluileProAl 25  
 |||||  
 Db 725 ATGTCACACCCCGCTTAAACGCACTGCGTGGCTGCTATTATTTAAAGAGATCCCTGC 784  
 |||||

Qy 25 aLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLys 42

```

Db      785  CCTGGATCGCAACTGCTCGACTGGCTGTGTCTGGAGGATTCATGACAAA 836
|||||
RESULT 8
LOCUS   AG192305
DEFINITION Pan troglodytes DNA, clone: RP43-068M07.T7, genomic survey
sequence.
ACCESSION AG192305
VERSION   AG192305.1 GI:45224481
KEYWORDS  Pan troglodytes (chimpanzee)
SOURCE    Pan troglodytes
ORGANISM  Pan troglodytes
REFERENCE 1
AUTHORS   Park,H., Kim,Y., Kim,S., Han,Y., Woo,T., Park,K., Eun,C.J.,
Hoon,S.T., Chu,M., Kim,H., Joo,S., Kim,C., Song,W. and Yoo,H.
TITLE     BAC end sequences of Library RP-43
JOURNAL   Unpublished
AUTHORS   Park,H., Kim,Y., Kim,S., Han,Y., Woo,T., Park,K., Eun,C.J.,
Hoon,S.T., Chu,M., Kim,H., Joo,S., Kim,C., Song,W. and Yoo,H.
TITLE     Direct Submission
JOURNAL   Submitted (07-JAN-2002) Hong-Seog Park, Korea Research Institute of
Bioscience and Biotechnology (KRIIB), Genome Research Center (GRC);
52, Oun-dong, Yuseong-gu, Daejeon 305-333, Korea
(E-mail:redstone@mail.kribb.re.kr, URL:http://phs.grc.kribb.re.kr/,
Tel:82-42-866-7181, Fax:82-42-860-4409)
COMMENT   Clones are derived from the chimpanzee BAC library RP-43 This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.
PRIMERS
Sequencing: T7
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES
Location/Qualifiers
1..440
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="RP43-068M07.T7"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RP-43 Chimpanzee Male BAC Library"
ORIGIN
Alignment Scores:
Pred. No.: 2,94e-07 Length: 440
Score: 160.00 Matches: 30
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 18.0% Indels: 0
DB: 14 Gaps: 0
US-10-718-311-16 (1-170) x AG192305 (1-440)
Qy 6 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluIleProAla 25
|||||
Db 349 ATGTACACCCCGCGTTAACGCAACTGCGTGGCTATTTGTAAGAGATCCCTGCC 408
|||||
Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeu 35
|||||
Db 409 CTGGATCCGCAACTGCTCGACTGGCTGTTG 438
|||||
RESULT 9
BZ551713 1110 bp DNA linear GSS 17-DEC-2002
LOCUS   pacs1-60_3287.x1 pacs1-60 Pseudomonas aeruginosa genomic clone
pacs1-60_3287, genomic survey sequence.
DEFINITION
ACCESSION BZ551713
VERSION   BZ551713.1 GI:27155321
KEYWORDS  Pseudomonas aeruginosa
SOURCE    Pseudomonas aeruginosa
ORGANISM  Pseudomonas aeruginosa
REFERENCE 1 (bases 1 to 1110)
AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE     Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL   J. Bacteriol. (2002) In press
COMMENT   Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
Location/Qualifiers
1..1110
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strains="1-60"
/db_xref="taxon:287"
/clone="pacs1-60_3287"
/clone_lib="pacs1-60"
/note="clinical isolate 1-60 Whole genomic shotgun
library."
ORIGIN
Alignment Scores:
Pred. No.: 1,42e-05 Length: 1110
Score: 150.00 Matches: 45
Percent Similarity: 37.7% Conservative: 12
Best Local Similarity: 29.8% Mismatches: 56
Query Match: 16.9% Indels: 38
DB: 11 Gaps: 3
US-10-718-311-16 (1-170) x BZ551713 (1-1110)
Qy 26 LeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThrLysArgPheGlu 45
|||||
Db 144 GTCGAACCCCTCTCTGAAGGCTGGCAGACCCTGCGGACGAC-----GAA 188
|||||
Qy 46 GlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 65
|||||
Db 189 TGCCAGGGGCTCGACGTC----- 206
|||||
Qy 66 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeu 85
|||||
Db 207 -----CCTACCGGACGTAGCGGCTGGGTCCGCGAGGTCTAC 242
|||||
Qy 86 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 105
|||||
Db 243 CTGCATGGCCACGACCGTCCCTGGGTGTGCGCCGAGCGTGGCGGCGCGCCCTG 302
|||||
Qy 106 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 125
|||||
Db 303 GAAGGCTCGGGCTTCGACCTGGCGCTGCTGGCACCGCTCGCTGGCGAGTTGCTGTTTC 362
|||||
Qy 126 ThrSerSerThrLeuThrArgAspPheIleGluIle----- 137
|||||
Db 363 AGCGACAGCGCCTTCAGCGCGGGGCCCATCGAAGTCTGCCGCTATCCGCGCGCCGCTCTG 422
|||||
Qy 138 -----GlyArgAspAlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGly 154
|||||
Db 423 CCCGCCGAGGTCCCGCGGAGGTTCTTGGGGCCGCTCGCTACGGTTTCCCGCGCGCG 482
|||||
Qy 155 LysProLeuLeuLeuThrGluLeuPheLeuPro 165
|||||

```

```

Db      483 CTCGGGTGCTGGTGGCGGAGTGACTACCG 515

RESULT 10
BZ564576/c
LOCUS   BZ564576             862 bp    DNA     linear     GSS 17-DEC-2002
DEFINITION pac82-164_4819.v2 pac82-164 Pseudomonas aeruginosa genomic clone
pac82-164_4819, genomic survey sequence.
ACCESSION BZ564576
VERSION   BZ564576.1  GI:27190626
KEYWORDS  GSS.
SOURCE    Pseudomonas aeruginosa
ORGANISM  Pseudomonas aeruginosa
          Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
          Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 862)
AUTHORS   Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
          Burns, J.L., Kaul, R. and Olsen, M.V.
          Whole-Genome-Sequence variation among multiple isolates of
          Pseudomonas aeruginosa library
JOURNAL   J. Bacteriol. (2002) In press
COMMENT   Contact: Chris K. Raymond
          Genome Center
          University of Washington
          Box 352145, Seattle, WA 98105-2145, USA
          Tel: 2062216954
          Fax: 2066857244
          Email: craymond@u.washington.edu
          Class: shotgun.
FEATURES  source
          location/Qualifiers
            1. 862
              /organism="Pseudomonas aeruginosa"
              /mol_type="genomic DNA"
              /strain="2-164"
              /db_xref="taxon:287"
              /clone="pac82-164_4819"
              /clone.lib="pac82-164"
              /note="clinical isolate 2-164 Whole genomic shotgun
              library."
ORIGIN
Alignment Scores:
Pred. No.: 0.000212      Length: 862
Score: 138.00           Matches: 45
Percent Similarity: 37.0%      Conservative: 11
Best Local Similarity: 30.3%    Mismatches: 56
Query Match: 15.6%            Indels: 38
DB: 11                    Gaps: 3

US-10-718-311-16 (1-170) x BZ564576 (1-862)

Qy      27 AspProGlnLeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGluGln 46
      : : : : :
Db      662 AACCCCTGCTGGAGCGCTGGCAGACCCTGGCGGACGAC-----GAATGC 618

Qy      47 GlnGlyLysThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGluIle 66
      : : : : :
Db      617 CAGGGGCTGACGTC----- 603

Qy      67 ProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluIleLeuLeu 86
      : : : : :
Db      602 -----CCTACCGCGAGTAGCGGCTGGGTCGGCGAGGTCCTACCTG 564

Qy      87 CysAlaAspGlyGluProTrpTrpLeuAlaGlyArgThrValValProValSerThrLeuSer 106
      : : : : :
Db      563 CATGCCACGACCGCTCCCTGAGTGTTCGCCGCGACGCGTGGCGGCGCGAGCGCCCTGGAA 504

Qy      107 GlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPheThr 126
      : : : : :
Db      503 GCTCGGGCTTCGACCTGGCGTCTCGCGACCCGCTCGCTGGGCGGAGTTGCTGTTTCAGC 444

Qy      127 SerSerThrLeuThrArgAspPheIleGluIle----- 137
      : : : : :
Db      443 GACAGCGCTTCGAGCGCGGGCCCATCGAAGTCTCGCGCTATCCGGCGCGCGTCTGCCC 384

```

---

```

Qy      138 -----GlyArgAspAlaGlyLeuTrpGlyArgArgSerArgLeuArgLeuSerGlyLys 155
      : : : : :
Db      383 GCCGAGGTCCGCGCGGAGGGTCTCTGGGGCGCTCGCTCACGGTTTTCGCCGCGCGCTC 324

Qy      156 ProLeuLeuLeuThrGluLeuPheLeuPro 165
      : : : : :
Db      323 GGGGTGCTGGTGGCGGAGGTGTACTACCG 294

RESULT 11
AJ863670/c
LOCUS   AJ863670             739 bp    DNA     linear     GSS 30-NOV-2005
DEFINITION Ralstonia solanacearum GSS, clone V789R, genomic survey sequence.
ACCESSION AJ863670
VERSION   AJ863670.1  GI:82937082
KEYWORDS  GSS; genome survey sequence.
SOURCE    Ralstonia solanacearum
ORGANISM  Ralstonia solanacearum
          Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
          Burkholderiaceae; Ralstonia.
REFERENCE 1 (bases 1 to 739)
AUTHORS   Munoz-Soriano, V., Arahall, D.R., Terol, J., Buaides, C., Perez-Perez, A.,
          Liop, P., Belmonte, U.C.F., Lopez, M. and Perez-Alonso, M.
          Random genome sequencing of Ralstonia solanacearum strain IVIA 1602
          and comparative analysis with strain GMI1000
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 739)
AUTHORS   Munoz-Soriano, V.
          Direct Submission
TITLE     Submitted (18-NOV-2004) Genetica, Universidad de Valencia, C/ Dr.
          Moliner, 50, Burjassot, Valencia 46100, SPAIN
JOURNAL
FEATURES  source
          location/Qualifiers
            1. 739
              /organism="Ralstonia solanacearum"
              /mol_type="genomic DNA"
              /strain="IVIA 1602"
              /db_xref="taxon:305"
              /clone="V789R"
ORIGIN
Alignment Scores:
Pred. No.: 0.028      Length: 739
Score: 118.00         Matches: 42
Percent Similarity: 38.9%      Conservative: 23
Best Local Similarity: 25.1%    Mismatches: 64
Query Match: 13.3%            Indels: 38
DB: 14                    Gaps: 3

US-10-718-311-16 (1-170) x AJ863670 (1-739)

Qy      30 LeuLeuAspTrpLeuLeuLeuGluAspSerMetThrLysArgPheGluGlnGlnGlyLys 49
      : : : : :
Db      599 CTGTTGATGTTGGTGAAGATCTTGAATTGCTGACGGCGGCTGCGTGGCGTTCTCTCC 540

Qy      50 ThrValSerValThrMetIleArgGluGlyPheValGluGlnAenGluIleProGluGlu 69
      : : : : :
Db      539 TCGTTCGCGTGAAGTCTCGGTGAG-----CAGCGGACC 504

Qy      70 LeuProLeu-----LeuProLysGluSerArgTyrTrpLeu 81
      : : : : :
Db      503 ATGCCGCTGAGCAGCAGTAGTGGCGTGTGCTGGGCTTGCCCGTGGCGGAGACGTTGGCG 444

Qy      82 ArgGluIleLeuLeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValPro 101
      : : : : :
Db      443 CGCAAGTCTCTGCTGATCTCGCAGCAAAACCCCGGTGGTCTTACGCACACACATCGTGAT 384

Qy      102 ValSerThrLeuSerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGly 121
      : : : : :
Db      383 CCGCGCAGCGTGGCGGGGACTGGCCGTTCTCTGAAGCGCTGGGACCCACCGCTCGGG 324

Qy      122 ArgTyrLeuPheThrSerSerThrLeuThrArgAspPheIleGlu----- 136
      : : : : :
Db      323 CATCGCTGTTTCGCGGATCCGCGGCTTGACACGGGCGGCTTCGAGTTTGGCCAGCTGGAT 264

```

```

QY 137 -----1141 bp DNA linear GSS 17-DEC-2002
Db 263 GTCCGGCATCGCTGCTCAGCGGGCGCGTGGCGCGCTGGGGCAGCGCGTGGGGGC 204
                                     :|:|:|
QY 140 AspAlaGlyLeuTrpGlyArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeu 159
Db 203 ATGGCGCGGTGCGCGCAGCGCGCTCGGTATTTTCGCGCGGTGCCAGCGCGATGCTGGTG 144
                                     :|:|:|
QY 160 ThrGluLeuPheLeuProAla 166
Db 143 ACAGAAAGTGTTCGCTGCA 123
                                     :|:|:|
RESULT 12
LOCUS BZ578851 1141 bp DNA linear GSS 17-DEC-2002
DEFINITION msh2 6034.x1 msh Pseudomonas aeruginosa genomic clone msh2_6034,
genomic survey sequence.
ACCESSION BZ578851
VERSION BZ578851.1 GI:27213912
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1141)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 20622216934
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
    Location/Qualifiers
        1..1141
            /organism="Pseudomonas aeruginosa"
            /mol_type="genomic DNA"
            /strain="MSH"
            /db_xref="taxon:287"
            /clone="msh2 6034"
            /clone_lib="msh"
            /note="Environmental isolate. Whole genomic shotgun
            library."
ORIGIN
Alignment Scores:
Pred. No.: 1.44 Length: 1141
Score: 105.00 Matches: 44
Percent Similarity: 36.2% Conservative: 24
Best Local Similarity: 23.4% Mismatches: 70
Query Match: 11.8% Indels: 50
DB: 12 Gaps: 6
US-10-718-311-16 (1-170) x BZ578851 (1-1141)
QY 22 GluIleProAlaLeuAspProGlnLeuLeuAspTrpLeuLeuGluAspSerMetThr 41
Db 466 CAATGTCCAGGTCTCCCGCCAGCTCTCGACTGGCTTTCGACGAGGGTCTCGTACC 525
QY 42 LysArgPheGluGlnGlnGlyLysThrValSerValThrMetIleArgGluGlyPheVal 61
Db 526 CGCGCGC-----CTGACCGTCTCGCGCAGCGCGGTTC 558
QY 62 GluGlnAsnGluIleProGluGluLeuProLeuLeu----- 73
Db 559 CGCGTGAACCCCTGCTGGNAGGTGGCAGACCCCTGCGCGCAGCAGCAATGCCAGGGGCTC 618

```

```

QY 74 -----ProLysGluSerArgTyrtTrpLeuArgGluIleLeuLeuCysAlaAspGlyGlu 91
Db 619 GAGTCCTCCCGGCGAGTAGCGGTCGGTCCGCGAGGTTACTCTGCATGCCACGACCGT 678
QY 92 ProTrpLeuAlaGlyArgThrValValProValSerThrLeuSerGlyPro----- 108
Db 679 CCTGGGTGNTCCCGCGCAGCGTGGCGGCGCGCGCTGAAGGGCTCGGCTTCGACT 738
QY 109 -----GluLeuAlaLeuGlnLysLeuGlyLysThrPro 119
Db 739 TGGCGCTGCTCGACCCCGCTGTTGGCGAGATTCTTTAGGAAGCGCCTTGAGCGCGGC 798
QY 120 LeuGlyArgTyrlLeuPheThrSerSerThrLeuThrArgAspPheIle-----GluIle 137
Db 799 CCATCAAGAGCTGGCGTATCCGCGCGCGCTTTGCCCGGCCAGTTCCCGCGGAGGTCCTCTG 858
QY 138 GlyArgAspAlaGlyLeuTrpGlyArgArgSerArgLeuArg-----Leu 152
Db 859 GGCCCGTCTTACGGTTTCCCGCGCGGCTGGGTGCTGTGGCGAAGGTACCTACCGGCGCTG 918
QY 153 SerGlyLysProLeuLeuLeuThr----- 160
Db 919 GGGACAGCCCGGAAATTGCCGACGATAAACCCCGCGTGAATTCGGAAGTCCCAAGTTTGG 978
QY 161 GluLeuPheLeuProAlaSerPro 168
Db 979 AACCTATCAAGCCCTTGTCCCT 1002
RESULT 13
LOCUS CL696699/753 bp DNA linear GSS 10-JUL-2004
DEFINITION PRI019d_B03_2 - PRI019d_BR (753), Mixed stage fosmid library of P.
aeruginosa var. California Pristionchus pacificus genomic, genomic
survey sequence.
ACCESSION CL696699
VERSION CL696699.1 GI:50218607
KEYWORDS Pristionchus pacificus
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 753)
AUTHORS Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
TITLE AppADB: an AcedB database for the nematode satellite organism
Pristionchus pacificus
JOURNAL Nucleic Acids Res. 32 (1), D421-D422 (2004)
PUBMED 14681447
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: raif.sommer@uebingen.mpg.de
This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.
Seq primer: T7
Class: fosmid ends.
FEATURES
    Location/Qualifiers
        1..753
            /organism="Pristionchus pacificus"
            /mol_type="genomic DNA"
            /strain="California"
            /db_xref="taxon:54126"
            /clone_lib="Mixed stage fosmid library of P. pacificus
            var. California"
            /note="Vector: pEpifos-5 Fosmid vector"
ORIGIN
Alignment Scores:
Pred. No.: 10.2 Length: 753
Score: 95.00 Matches: 19
Percent Similarity: 100.0% Conservative: 0

```



Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 10.7% Indels: 0  
 DB: 13 Gaps: 0

US-10-718-311-16 (1-170) x CL696699 (1-753)

QY 152 LeuSerGlyLeuProLeuLeuThrGluLeuPheLeuProAlaSerProLeuTyr 170  
 DB 751 TTAAGCGGTAAACCGCTGTGCTACAGAACTGTTTTTACC GGCGTCACCGTTGTAC 695

#### RESULT 14

LOCUS BG799182 591 bp mRNA linear EST 21-MAY-2001  
 DEFINITION f3p1c10.y1 zebrafish gridded kidney Danio rerio cDNA clone  
 IMAGE:4744722 5' similar to TR:Q9Y4G8 Q9Y4G8 KIAA0313 PROTEIN. ;,  
 mRNA sequence.

#### ACCESSION

VERSION BG799182.1 GI:14163514

#### KEYWORDS

SOURCE EST.

#### ORGANISM

Danio rerio (zebrafish)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.

#### REFERENCE

AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M.,  
 Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,  
 Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,  
 Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,  
 Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,  
 Waterston,R. and Wilson,R.  
 WASHU Zebrafish EST Project 1998

#### TITLE

WASHU Zebrafish EST Project 1998

#### JOURNAL

Unpublished (1998)

#### COMMENT

Contact: Stephen L. Johnson  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: zbrafish@watson.wustl.edu  
 cDNA Library Preparation: Leonard Ira Zon DNA Sequencing by:  
 Washington University Genome Sequencing Center Clone distribution:  
 Genome Systems, St. Louis, Missouri (web address:  
 www.genomesystems.com) (email contact: info@genomesystems.com) and  
 Research Genetics, Huntsville, Alabama (web address:  
 www.resgen.com) (email contact: info@resgen.com) and  
 RessourcenzentrumPrimarDatenbank, Berlin, Germany (web address:  
 www.rzpd.de)

High quality sequence stop: 473.

#### FEATURES

##### source

1..581  
 Location/Qualifiers  
 /organism="Danio rerio"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7955"  
 /clone="IMAGE:4744722"  
 /sex="mixed"  
 /tissue\_type="kidney pooled from 300 wild type adults"  
 /lab\_host="XLOLR"  
 /clone\_lib="zebrafish gridded kidney"  
 /note="Organ: kidney; Vector: pBK-CMV; Site 1: EcoRI;  
 Site 2: XhoI; Oligo dt cDNA library constructed from mRNA  
 pooled from pooled kidney tissue from 300 adult  
 zebrafish."

#### ORIGIN

##### Alignment Scores:

Pred. No.: 11.7 Length: 581  
 Score: 93.00 Matches: 28  
 Percent Similarity: 45.4% Conservative: 26  
 Best Local Similarity: 27.0% Mismatches: 43  
 Query Match: 10.5% Indels: 34  
 DB: 2 Gaps: 7

US-10-718-311-16 (1-170) x BG799182 (1-581)

QY 7 SerHisProAlaLeuThr-----GlnLeuArgAlaLeuArgTyrCysLysGluLeuPro 24  
 DB 117 AGTAACCCCTGACCTGGCACAGGCGCAGACGCATCATCGACTACAGTACACGACCA 176  
 QY 25 AlaLeuAspProGlnLeuLeu-----AspTrpLeuLeu 35  
 DB 177 GAGTTGCAAGACCAGGTGCTCGCATTTTCAAGGCAGATCAGCAAGCCGTTACTACTG 236  
 QY 36 LeuGluAspSerMetThrLysArg-----PheGluGlnGlnGlyLys 49  
 DB 237 GCCAATAGACACACACAGCCAGAGAGCTGCCAATCTAGCCATAAAGAGTTTGCCTTG 296  
 QY 50 ThrValSer-----ValThrMetIleArgGluGlyPheVal 61  
 DB 297 TCTCGAGTCCGGAAGCTTTTTCGCTCTGTGAAGTTTCAGTCACACAGGAAGGAGTTATC 356  
 QY 62 GluGlnAsnGluIleProGluGluLeuProLeuLeuProLys-----GluSer 77  
 DB 357 AAGCAGAGACACTGCTGTGATCAACTGTCAAACTGGCTGACAGATCCAAGTGAAGTGCC 416  
 QY 78 ArgTyrTrpLeuArgGluIleLeu-----LeuCysAlaAspGlyGluProTrp 93  
 DB 417 AGGTACTATCTAAGAGCAATATGMAACCGAGACGCTGTCTCGATGTGGAAGCACTG 476  
 QY 94 ---LeuAlaGlyArgThrValProValSerThrLeuSerGlyProGluLeuAlaLeu 112  
 DB 477 GAGATGCAGAGAGAATCTGTGTCGCCCTGTTGTCTTTGAGCTCAATGGAGATAGCCAAT 536  
 QY 113 Gln 113  
 DB 537 CAG 539

#### RESULT 15

##### LOCUS

CR861280

##### DEFINITION

Pongo pygmaeus mRNA; cDNA DKFZp459N063 (from clone DKFZp459N063).

##### ACCESSION

CR861280

##### VERSION

CR861280.1 GI:55733331

##### KEYWORDS

HTC.

##### SOURCE

Pongo pygmaeus (orangutan)

##### ORGANISM

Pongo pygmaeus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Pongo.

1 (bases 1 to 4527)

Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Oeanger,A.,

Fobo,G., Han,M. and Wiemann,S.

The German CDNA Consortium

Direct Submission

Submitted (12-NOV-2004) MIPS, Ingolstaedter Landstr.1, D-85764

Neuherberg, GERMANY

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

sequenced by Agowa (Berlin/Germany) within the cDNA sequencing

consortium of the German Genome Project.

This clone (DKFZp459N063) is available at the RZPD Deutsches

Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.

Please contact RZPD for ordering:

http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp459N063

Further information about the clone and the sequencing project is

available at http://mips.gsf.de/projects/cdna/.

##### FEATURES

##### Location/Qualifiers

1..4527

/organism="Pongo pygmaeus"

/mol\_type="mRNA"

/db\_xref="taxon:9600"

/clone="DKFZp459N063"

/tissue\_type="cortex"

/clone\_lib="459 (synonym: pcor1). Vector pSport1\_Sfi; host

DH10B; sites SfilA + SfilB"

/dev\_stage="adult"

/note="ATPase, H+ transporting, lysosomal accessory



gene protein precursor (Homo sapiens) "

1..4527

/gene="DKFZp459N063"

17..367

/gene="DKFZp459N063"

/codon\_start=1

/product="hypothetical protein"

/protein\_id="CAH93348.1"

/db\_xref="GI:55733332"

/translation="MMAATATARVRMGPRCAQALWEMPMLPVFLSLAAAVVAARAAA

VAEQVPLVLMSSDRDLNAPADTHEGHITSLQLSTYLDPALELGPRLVLLFLQDKV

RPQPTLPQSSGS"

# ORIGIN

## Alignment Scores:

Pred. No.: 252 Length: 4527  
 Score: 92.50 Matches: 48  
 Percent Similarity: 35.1% Conservative: 23  
 Best Local Similarity: 23.8% Mismatches: 76  
 Query Match: 10.4% Indels: 55  
 DB: 6 Gaps: 8

US-10-718-311-16 (1-170) x CR861280 (1-4527)

QY 4 TrpHisMetSerHisProAlaLeuThrGlnLeuArgAlaLeu----- 17  
 Db TGGAGGGCAAGGTGCGCAGCTGCTACAAAAACAGCCAGTATCACCTGTGATCCTCTCC 3270  
 QY 18 -----ArgTyrCysLysGluIleProAlaLeuAspProGlnLeuLeu----- 31  
 Db TGTGAGTTACACGACACCGCTCCCGGATCTCTTCTGGGCCCAAACTTCTCTGTGGC 3330  
 QY 31 ----- 31  
 Db GTACAGGACCAAGTGGAGGACCTGACTCCCTCACCCTTGGGTGTCAGAGCTCAACCT 3390  
 QY 32 AspTrpLeuLeuGluAspSerMetThrLysArgPheGluGlnGlnGlyLysThrVal 51  
 Db GACTGGCTCTTCTGGATGACTCCTTTGCCAGGCTCTCACTGACCTATGAACGACTCTT 3450  
 QY 52 SerValThrMetIleArgGluGlyPheValGluGlnAsnGluIleProGluGluLeuPro 71  
 Db TGGTACCACAGTGACATTCAAGGTCTATTCTGGCCCAACCGCCTCTACCCAGTATCTGCCCG 3510  
 QY 72 LeuLeu-----ProLysGluSerArgTyrTrpLeuArgGluIleLeuLeu 86  
 Db GCACGTGGTTTACATGGAGCGCTCGAAGTCCACAGCAATGGCTCGTCGCTACTTCAA 3570  
 QY 87 CysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeuSer 106  
 Db TGCTTCCCAAGGTCAC-----AGGGCCACAGCATCTACTCTTCCACTGGGAGTATGTACAG 3624  
 QY 107 GlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPheThr 126  
 Db CAGCCTGAG-----CAAGAAGGGGAGTCTCTCTGTCGGCCCGCAC----- 3663  
 QY 127 SerSerThrLeuThrArgAspPheIleGluIleGlyArgAspAlaGlyLeuTrpGlyArg 146  
 Db GCAGCCCTCCTCTCTGGCAGATGATGCTTCAGGACTTCCAGATCCAGGCTTT----- 3714  
 QY 147 ArgSerArgLeuLeuSerGlyLysProLeuLeu-----LeuThrGluLeuPhe 163  
 Db -----CAACGTAATGGGGAGAGAGTTCCTTACGCCAGGAGCTGTGCCAGCTTCTT 3765  
 QY 164 LeuPro 165  
 Db CTCCCC 3771

Search completed: May 30, 2006, 04:08:17  
 Job time : 4396.04 secs

**THIS PAGE BLANK (USPTO)**

XX WPI; 1995-186908/25.  
 DR P-PSDB; AAR74742.  
 XX Transgenic plants with increased content of resistance factor - contg.  
 PT bacterial DNA coding for enzyme catalysing prodn. of resistance factor,  
 PT e.g. antiviral p-hydroxy-benzoic acid.  
 XX Claim 3; Page 5; 7pp; German.  
 XX Transgenic plants containing a bacterial gene which encodes an enzyme  
 CC able to catalyse prodn. of an antiviral, bactericidal, fungicidal or  
 CC insecticidal factor are new. A preferred gene is ubiC from E.coli  
 CC (AA92409) which codes for chorismate-pyruvate lyase (AAR74742). The  
 CC lyase catalyses conversion of chorismate to p-hydroxybenzoic acid and  
 CC transgenic tobacco plants which express the active enzyme are resistant  
 CC to tobacco mosaic virus  
 XX Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;  
 SQ

Alignment Scores:  
 Pred. No.: 6,52e-89 Length: 495  
 Score: 854.00 Matches: 165  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 100.0% Indels: 0  
 DB: 2 Gaps: 0

US-10-718-311-4 (1-165) x AAQ92409 (1-495)

QY 1 MetSerHisProAlaLeuThrGlnLeuArgAlaLeuArgTyrCysLysGluLeuProAla 20  
 DB 1 ATGTACACCCCGCGTAAACGACATGCTGCGTGGCTGCTGATTTGTAAGAGATCCCTGCC 60  
 QY 21 LeuAspProGlnLeuLeuAspTropLeuLeuGluAspSerMetThrLysArgPheGlu 40  
 DB 61 CTGGATCCCACTGCTCGACTGGCTGCTGCTGGAGATTCATGACAAACGTTTGA 120  
 QY 41 GlnGlnGlyLeuThrValSerValThrMetIleArgGluGlyPheValGluGlnAsnGlu 60  
 DB 121 CAGCAGGGAACCGGTAGCGTACGATGATCCGCGAGGGTTTGTCCGACGATGAA 180  
 QY 61 IleProGluGluLeuProLeuLeuProLysGluSerArgTyrTrpLeuArgGluLeuLeu 80  
 DB 181 ATCCCGAAGAACTGCCGCTGCTGCCGAAAGAGTCTCGTTACTGCTTACGTAATTTTG 240  
 QY 81 LeuCysAlaAspGlyGluProTrpLeuAlaGlyArgThrValValProValSerThrLeu 100  
 DB 241 TTATGTGCCGATGGTGAACCGTGGCTTGCCTGCGTACCGTCTCTGTGCAACGTTA 300  
 QY 101 SerGlyProGluLeuAlaLeuGlnLysLeuGlyLysThrProLeuGlyArgTyrLeuPhe 120  
 DB 301 ACCGGCGCGAGCTGGCGTTACAAANTTGGTAAACCCGTTAGGACGCTATCTGTTTC 360  
 QY 121 ThrSerSerThrLeuThrArgAspPheIleGluLeuGlyArgAspAlaGlyLeuTrpGly 140  
 DB 361 ACATCATCGACATTAACCCGGCACTTATTGATAGAGCCGATGCCGGGTGTGGGGG 420  
 QY 141 ArgArgSerArgLeuArgLeuSerGlyLysProLeuLeuLeuThrCluLeuPheLeuPro 160  
 DB 421 CGACGTCCCGCTGCGATTAAGCGGTAAACCGCTGTTGCTAACGAACTGTTTACCG 480

QY 161 AlaSerProLeuTyr 165  
 DB 481 GGTACCGTGTAC 495

RESULT 2  
 ID ABA91837  
 XX ABA91837 standard; DNA; 495 BP.  
 AC ABA91837;  
 XX 15-MAY-2002 (first entry)

XX Escherichia coli chorismate pyruvate lyase gene coding region.  
 DE  
 XX Chorismate pyruvate lyase; CPL; enzyme; p-hydroxybenzoic acid;  
 KW transgenic plant; ubiC; gene; ds.  
 XX Escherichia coli.  
 OS  
 XX Location/Qualifiers  
 FH 1..495  
 CDS /tag= a  
 FT /partial  
 FT /product= "chorismate pyruvate lyase"  
 FT /note= "the CDS does not include a stop codon"  
 XX WO200194607-A2.  
 PN  
 XX 13-DEC-2001.  
 PD  
 XX 22-MAY-2001; 2001WO-US016661.  
 PF  
 XX 02-JUN-2000; 2000US-0209854P.  
 PR  
 XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
 PA  
 XX Meyer K, Van Dyk DE, Viitanen PV;  
 PI WPI; 2002-226795/28.  
 XX P-PSDB; AAM50958.  
 DR  
 XX Producing para-hydroxy benzoic acid in green plant, comprises expression  
 PT of unique expression cassette containing gene encoding chorismate  
 PT pyruvate lyase operably linked to specific chloroplast targeting  
 PT sequence.  
 PT  
 XX Claim 4; Page 55; 60pp; English.  
 PS  
 XX The present sequence is that of the coding region of the Escherichia coli  
 CC strain W3110 ubiC gene in expression construct pET24a. The gene encodes  
 CC chorismate pyruvate lyase (CPL, see AAM50958). It was obtained by PCR  
 CC amplification of strain W3110 genomic DNA using primers (see ABA91839-40)  
 CC based on the published E. coli ubiC gene. The CPL open reading frame is  
 CC used in an expression cassette designed for the high-level production of  
 CC p-hydroxybenzoic acid (pHBA) in green plants. The expression cassette  
 CC comprises the CPL coding sequence operably linked to a promoter capable  
 CC of driving protein expression in higher plants. The cassette also has a  
 CC sequence encoding a chloroplast transit peptide, its natural cleavage  
 CC site, and a small portion of a transit peptide donor protein fused to the  
 CC N-terminus of CPL. The chloroplast targeting sequence targets the foreign  
 CC protein to the chloroplast and aids in its uptake into the organelle. The  
 CC cleavage site is unique to the transit peptide, and cleavage of the  
 CC chimeric protein encoded by the cassette at this site releases a novel  
 CC polypeptide that has full enzyme activity, comprising the mature CPL  
 CC enzyme and a small portion of transit peptide donor (see AAM50961). A  
 CC plant comprising the CPL expression cassette is claimed, and may be  
 CC soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat,  
 CC barley, oats, sorghum, rice, Arabidopsis, sugarcane, canola,  
 CC millet, bean, pea, yve, flax or a forage grass. pHBA is a monomeric  
 CC component of liquid crystal polymers which have application in the  
 CC automotive, electrical and other industries  
 XX  
 SQ Sequence 495 BP; 112 A; 122 C; 141 G; 120 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 6,52e-89 Length: 495  
 Score: 854.00 Matches: 165  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 100.0% Indels: 0  
 DB: 6 Gaps: 0

US-10-718-311-4 (1-165) x ABA91837 (1-495)

**THIS PAGE BLANK (USPTO)**